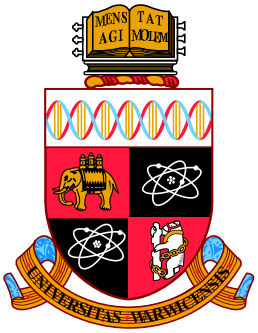
Condition-dependent transcriptional landscape of *Campylobacter jejuni*



Jenna Hoi-Yan Lam

Doctoral thesis

Submitted for the degree of Doctor of Philosophy in food security

Supervisor: Dr. Chrystala Constantinidou

Co-Supervisors: Dr. Mohammad Tauqeer Alam and Dr. Emma Denham

School of Life Sciences

University of Warwick

September 2019

Contents

List of Tables

List of Figures

Acknowledgements

"Because I knew you, I have been changed for good"

-Elphaba, (For Good ♪), *Wicked*, Stephen Schwartz

First of all, I really appreciate the concept of having acknowledgements as part of the thesis. The PhD journey can be a very hard and difficult time, but also a life-changing experience and the people I have met on the way, good and bad, have shaped me into who I am today. For that I am very grateful and would like to express my sincerest gratitude for those who have supported me throughout these years. That's why my first quote is from the song "For Good" from the musical *Wicked*. I would just like to add that the idea of putting quotes in acknowledgements was actually inspired by Joe's thesis!. Lets see if it becomes a trend. The entire song "For Good" really resonates with how I feel in this snapshot of a moment right at the very end of the PhD journey. It is a very inspirational and heart-warming song and I highly recommend anyone to give it a listen. But the quote "Because I knew you, I have been changed for good" applies to everyone whom I have met past and present in my life, but especially those in the past four years who have been an integral part of my journey.

Thank you wholeheartedly to my supervisor Dr. Chrystala Constantinidou, for without whom this project would not have commenced. For her patience, humour, excitement, kindness, dealing with my night-owl schedule/routine, and truly being a supportive supervisor who has always been on my side. I have been very lucky and am very grateful for all her ideas, guidance, and mentoring throughout. This project is just as much her effort as mine. Thank you for always being so approachable and understanding, I have learned a lot from you :). I would also like to thank my second supervisor Dr. Emma Denham for her generosity and expertise whilst she was at Warwick, and for treating me as one of her own students. Last but not least, I would like to thank Dr. Mohammad Tauqeer Alam who has been very kind in taking up the role of second supervisor in my final year of research, and on top of helping me immensely with data analysis, has given me very thoughtful advice that sometimes gives a new perspective to life. I hope one day to be just as philosophical and knowledgeable about science and life.

I would also like to extend my gratitude to my advisory panel Dr. Andrew Millard, Dr. David Scanlan, and Dr. Irene Stefanini for all the useful advice throughout the project and for always putting my best interests

at heart. In particular, Irene who was very kind and generous to take out her time to help me solve issues with my data.

Now onto everyone else in the formerly known Microbiology Infection Unit. As there are so many of you past and present, I am sorry if I have not included your name but thank you for being part of the lab and for making the group such a sociable and relaxed environment, to have fun together with (this includes PIs too). I will definitely miss the lunch breaks, the odd coffee breaks that I do attend, the random Friday fry-ups, which I rarely attend because of how early it is, and all the Friday pubs, parties, and Christmas meals. All the weird and wonderful traditions, which hopefully will still be up-held in the future and maybe evolve. All the hilarious conversations in the office whilst everyone was procrastinating. To have such a good lab environment is unfortunately not as common as we'd like, and I will never take it for granted <3.

To all the postdocs in the lab former and present. Thank you Dr. Alexia Hapeshi for basically being the proxy postdoc for everyone in the lab, especially for my group, which doesn't have one. Your knowledge and ideas have helped me so much and you really do inspire us. Thank you to Dr. Joseph Healey for taking his time to help me with command-line and latex and sharing all his knowledge. Thank you to Dr. Sébastien Raguideau for being patient with teaching me Python even though I didn't use it much in the end. But also the conversations we had about anime, playing Gloomhaven, and basically all things geeky, which I really appreciate! To Dr. Josie Mckeown, for all the late night conversations we had in the lab as a fellow night owl and knowledge about RNA, to Dr. Arnaud Kengmo Tchoupa for always giving me deep insights into life and having a very unique sense of humour and also sometimes letting me share some of his reagents, to Dr. Sandra Bedarida for sharing sequencing reagents and knowledge and also helping me buy various things while I was writing!, to Dr Nina Luhmann and Dr. Blessing Anonye for being so kind and checking up on me during the PhD. Special mention to Dr. Povilas Sazinas, for helping me out with his toRNAdo tool but also our shared interest in mobile games.

To my fellow PhD students. Firstly, I would like to thank you Swati Mahapatra for who we started ("officially" for me anyway) our PhDs together and for being my uni wife. Although, we cannot end it together but we started together the journey together and it will always be forever memorable. For all the shared pain we went through doing research, the conferences we attended together, and mutual support as PhD students, I will always be forever grateful that I have someone to rant and talk to. Thank you to Leo Souliotis, for being so generous in taking his time to help us "biology" students with writing scripts and for our mutual love of music and karaoke. Thank you to Shathviga Manoharan for being the sweetest, to always check up on me during the write-up period and being my hilarious alternative asian sister. Special mention to Thomas Brooker who helped me break out of my shell when I first joined the lab even though sometimes you were a bit mean haha.

To the Constantinidou lab group members. My most sincere gratitude to Dr. Emily Stoakes who til this day is still so kind and caring and for previously being the only senior in the group. Thank you for being so

understanding and for all the advice, mentoring, and assurance you have given me. Thank you to Stephen Li for tolerating and helping me with all the R scripts and for reaching things too high up for me to take haha, and also having a shared love for anime. Thank you to Dagnija Tupina for being such an integral member of the lab and also liking anime and 9gag. (Also Shath but you're already mentioned and you're in Nick's group now :p).

I would like to extend my utmost gratitude to my housemate Paloma Perez Galvan with whom I have lived together for three years. I am so grateful to live with someone who also likes watching Korean dramas and anime, listening to Kpop, and going to the gym together. Thank you for also listening to all my woes and worries and for sharing stories about our PhDs and the spiders in our rooms! I would also like to thank you my friend Dr. Almutasem Saléh or Al as he likes to be called for being my friend since the beginning of undergrad so we basically passed the 7 year mark! and for sharing his experiences as a PhD student and advice. I would also like to thank his brother Huthaifah Saléh or H for also taking the time to listen to my PhD woes. In addition, I would like to thank you my housemates from first year in Claycroft who I only met for one year but we made good memories together: Yuyang, Sherry, and Katherine. Also thank you to the people who I have met in Warwick ABACUS for making me feel like I actually belonged somewhere.

I would like to thank everyone in the 2015 MIBTP cohort for starting and finishing this PhD journey together. To always have people in the same boat, going through the same deadlines is always comforting. I would like to have special mentions to Magdalena Karlikowska who was my first friend on the course and like my total opposite, although we have shared traits of leaving things til last minute, but I guess opposites attract haha. Also to Julia Lipecki and Chris O'Grady for always being so funny, and Robert Millar and Olivia Nippe for being so sweet.

Most importantly and finally, I thank with all my heart my family or as we like to refer to as famalam for all their endless support and love. Without all their help, this PhD would be near impossible if not extremely difficult to complete. I thank my parents, and my younger sister and younger brother for tirelessly and unconditionally helping me to move accommodation, bringing me homemade food, encouraging me, believing in me, but also never giving me any pressure to succeed and letting me know that my mental health is a priority. I love my family and they love me dearly too. That is why I would like to end on the quote from the anime *Mob Psycho*, which I also highly recommend, that "People need other people". Whether it is friends and/or family, or colleagues and supervisors, we all succeed by helping each other out and caring for one another. I apologise for the three page acknowledgements.

"People need other people"

-Mob, *Mob Psycho 100*, ONE

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Chapter 4: The R script for extracting iron-repressed TPM values TSS_iron.R was written with the help of Dr. Mohammad Tauqeer Alam.
- Chapter 4: All other R scripts used for this chapter were written with the support of Stephen Li unless otherwise specified.
- Chapter 5: The original Python script for filtering coverageBed read counts PrepareDESeq_gff3_locus.py was written by Dr. Richard Brown.
- Chapter 5: The Python script for de-multiplexing RNAtag-Seq fastq files Example_tile.py was written by Leonidas Souliotis.
- Chapter 5: The R script for combining read counts combining_read_counts.R was written with the help of Swati Mahapatra.
- Chapter 5: The R script for automating pairwise comparisons of samples in DESeq2 DESEQ2analysis_automated_ 22_conditions.R was written with the help of Dr. Joseph Richard John Healey.
- Chapter 5: The R script for grouped comparisons in DESeq2 master_comparisons.R was written with the help of Dr. Mohammad Tauqeer Alam.

Abstract

Campylobacter spp. are ubiquitous in the environment and can infect a plethora of animals, *Campylobacter jejuni* being the most predominant species infecting humans. One striking characteristic of this pathogen is its supposedly commensal status in avian species as opposed to an infectious agent in other animals. The main route of transmission to the human host is via consumption of undercooked poultry in developed countries, although there are other ways to become exposed such as through contact with infected animals or ingestion of contaminated water.

The sequenced reference strain NCTC 11168 revealed many features including a small tightly packed genome with three sigma factors and a limited number of global regulators. Furthermore, *C. jejuni* lacks many transcriptional factors present in other bacteria for rapid alteration of gene expression. Alternative forms of regulation such as small regulatory RNAs are not well characterised in *C. jejuni*, but may play a role in fine-tuning expression under different stresses. Studies in the literature have investigated the transcriptome of NCTC 11168 under standard growth or specific stress conditions, but this is the first time the transcriptional landscape of NCTC 11168 was determined under a collection of 22 host-relevant and transmission conditions, using a combination of high-throughput sequencing methodologies: RNAtag-Seq and Cappable-seq.

A highly reproducible data set representing the condition-dependent transcriptional landscape of NCTC 11168 was generated and assessed. This extensive data set includes gene expression profiles of individual conditions and the primary transcriptome of all combined conditions. A considerable repertoire of transcriptional start sites (TSS) were identified that corroborate with published data, with a significant number that were novel condition-dependent TSS and mainly internal. Preliminary analysis also revealed a subset of non-coding RNAs with an associated TSS. The behaviour of gene expression profiles was analysed and enriched for KEGG pathways with a focus on the behaviour of the stringent response. The data obtained from this project provides a good resource for the *Campylobacter* community, and is also a basis for future work to help decipher the regulatory systems of *C. jejuni* and give clues to how it can successfully act as a foodborne-pathogen.

List of Abbreviations

°C	degrees Celsius
g	g-force
σ	sigma
μg	microgram
μL	microlitre
μM	micromolar
(p)ppGpp	guanosine 5'-diphosphate 3'-diphosphate (ppGpp)/guanosine 5'triphosphate 3'-diphosphate (pppGpp)
A	Adenine
ABC	ATP-binding cassette
ACP	acyl carrier protein
ACU	Atmospheric control unit
BAM	Binary Alignment Map
bcl	binary base call
BLAST	Basic local alignment search tool
bp	base pair
C	Cytosine
CCVs	*Campylobacter* containing vacuoles
cDNA	complementary DNA
CFU	colony-forming unit
Cia	*Campylobacter* invasion antigens
CLIMB	Cloud Infrastructure for Microbial Bioinformatics
CO_2	carbon dioxide
CsrA	Carbon storage regulator/Regulator of secondary metabolism
csv	comma separated value
DNA	deoxyribonucleic acids
DNase	deoxyribonuclease
dRNA-seq	differential RNA-sequencing
DTB-GTP	3'-desthiobiotin-TEG-guanosine 5' triphosphate
EDTA	ethylenediaminetetraacetic
fdr	false discovery rate

Fur	ferric uptake regulator
G	Guanine
GBS	Guillain Barré syndrome
GFF	general feature format
GSNO	S-Nitrosoglutathione
HCl	hydrogen chloride
Hfq	HF-I protein
HTS	high-throughput sequencing
kb	kilobases
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	litre
log	logarithmic
LOS	lipooligosaccharide
m	metre
M	molar
Mb	megabase
MEMα	minimum essential medium alpha modification
MEME	Multiple Em for Motif Elicitation
mg	milligram
Mg^{2+}	magnesium
$MgCl_2$	magnesium chloride
MH	Mueller-Hinton
MIC	minimum inhibitory concentration
mL	millilitre
mM	millimolar
mRNA	messenger RNA
NaCl	sodium chloride
NaN_3	sodium azide
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology
ncRNA	non-coding RNA
NCTC	National Collection of Type Cultures

NEB	New England Biolabs
ng	nanograms
nM	nanomolar
nt	nucleotide
O_2	oxygen
OD	optical density
OMV	outer membrane vesicles
orf	open reading frame
P	monophosphate
PBS	phosphate buffered saline
PCA	Principle Component Analysis
PCR	polymerase chain reaction
PerR	peroxide resistance regulator
PNPase	polynucleotide phosphorylase
poly-P	inorganic polyphosphates
PPP	triphosphate
RBS	ribsome binding site
RIN	RNA integrity number
rlog	regularized log
RNA	ribonucleic acids
RNA-seq	RNA-sequencing
RNase	ribonuclease
RPK	reads per kilobase
rpm	revolutions per minute
rRNA	ribosomal RNA
RRS	relative read score
s	seconds
SAM	Sequence Alignment Map
SCID	severe combined immune deficiency
SDS	sodium dodecyl sulphate
SMRT	Single Molecule, Real-Time
SNPs	single nucleotide polymorphisms

STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
T	Thymine
T3SS	type III secretion system
TAP	tobacco alkaline phosphatase
TCA	tricarboxylic acid
TE	10 mM Tris, 1 mM EDTA
TPM	transcripts per million
tRNA	transfer RNA
TSS	transcriptional start site
UTR	untranslated region
v/v	volume/volume
VAIN	Variable atmosphere incubator
VBNC	viable but non-culturable
VCE	Vaccinia capping enzyme
w/v	weight/volume

Chapter 1

Introduction

1.1 Reclassification of *Campylobacteraceae*

For many years, the Epsilon subdivision of the Proteobacteria i.e. Epsilonproteobacteria, was given class status encompassing two families: *Campylobacteraceae* and *Helicobacteraceae* (Waite *et al.*, 2018). These two families are well known for their pathogenic species *Campylobacter jejuni* and *Helicobacter pylori* respectively. In particular, *Campylobacteraceae* is the largest and most diverse of the two consisting of the three genera: *Campylobacter*, *Arcobacter*, and *Sulfurospirillum*, which include environmental as well as commensal and pathogenic species, occupying a broad range of roles and niches (Lastovica *et al.*, 2014). Until 1963, *Campylobacter* spp. were actually classified as part of the *Vibrio* genus due to similarities in cell morphology (Véron and Chatelain, 1973). Since their G/C content differed greatly as well as other phenotypic features, a new genus was proposed by Sébald and Véron (1963) named as *Campylobacter*. This was the beginning of many changes to classifications in *Campylobacteraceae* and especially in *Campylobacter*, as new taxonomic methods became more robust and streamlined (Lastovica *et al.*, 2014).

Epsilonproteobacteria have recently undergone even more scrutiny with proposals by Waite *et al.* (2018) to reassign Epsilonproteobacteria to a new phylum called Epsilonbacteraeota, based on concatenated protein markers and ribosomal gene sequence phylogeny (Waite *et al.*, 2018). These subsequent changes have also affected *Campylobacteraceae*; *Arcobacter* and *Sulfurospirillum* are now their own respective families and the *Campylobacter* genus has been further subdivided as seen in Figure 1.1 (Waite *et al.*, 2018). In this project, the main focus is on the Gram-negative host-associated pathogenic species *C. jejuni* from the *Campylobacter* genus, known for its clinical burden.

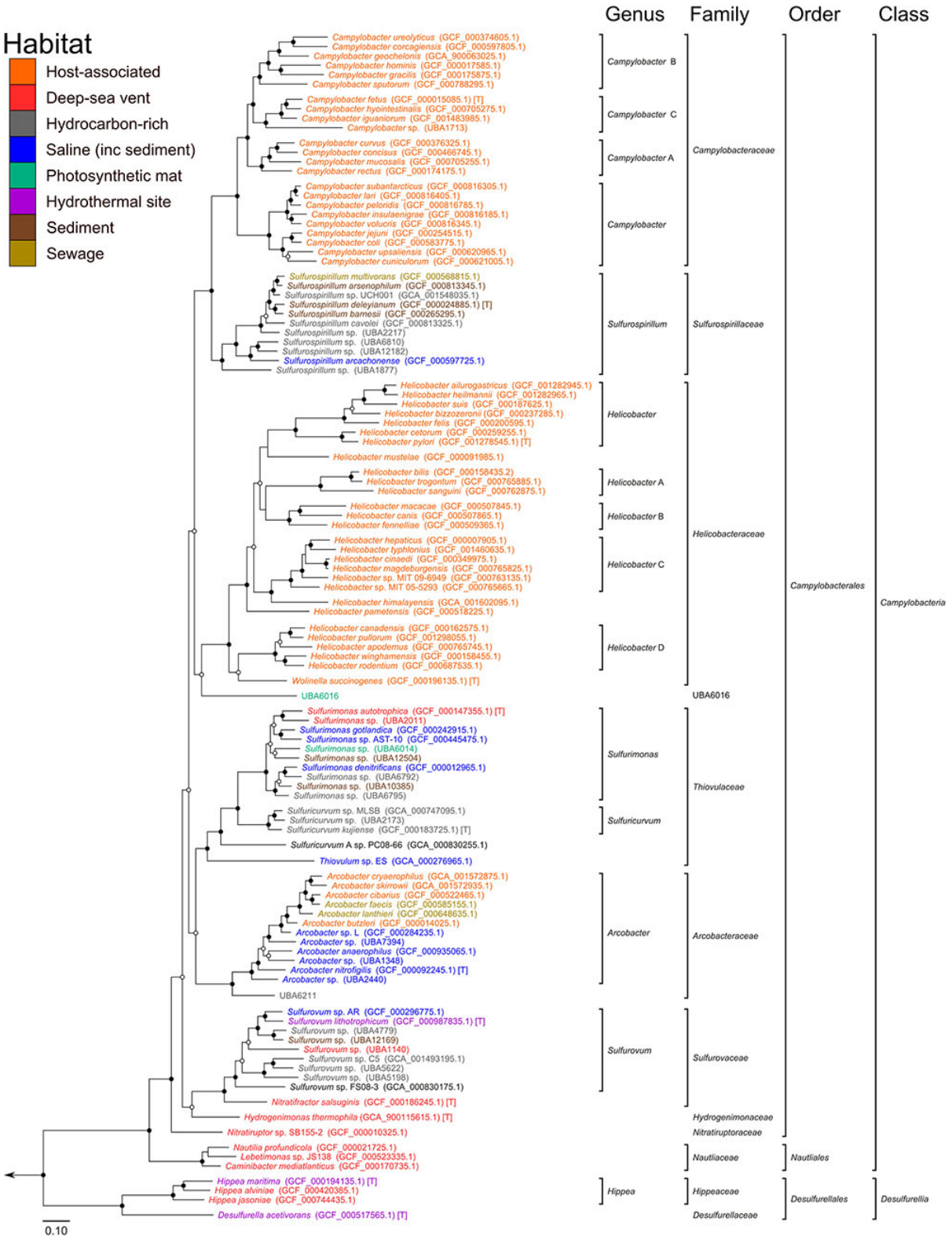


Figure 1.1: Maximum likelihood tree from concatenated protein markers of the newly reclassified phylum of Epsilonbacteraeota. Taken from Waite *et al.* (2018).

1.2 Background of *Campylobacter jejuni*

The Gram-negative, non-spore forming, microaerophilic and capnophilic, zoonotic bacterium, *C. jejuni* is the leading cause of bacterial gastroenteritis in developed countries and is endemic in low-income areas with poor access to infrastructure (WHO, 2018). *C. jejuni* is typically corkscrew shaped and is highly motile, possessing uni-polar or bi-polar flagella located at either end of the cell that are important for colonisation (Figure 1.2). The morphology of *C. jejuni* changes during its growth; actively growing cells are S or spiral shaped whereas old inactive cells are coccoid shaped (Ng *et al.*, 1985). The latter is an indication of the viable but non-culturable (VBNC) state (Rollins and Colwell, 1986). *Campylobacter* cells are exceptionally small around 0.2-0.8 μM wide and 0.5-5 μM long as seen in Figure 1.3. *C. jejuni* is a mesophile with reported growth at a minimum of 30 °C and maximum 45 °C (Skirrow and Benjamin, 1980). This range covers the human body temperature at 37 °C and avian body temperatures including chickens at 42 °C. The narrow temperature range and microaerophilic growth conditions as well as nutrient requirements render it fastidious to maintain in the laboratory.

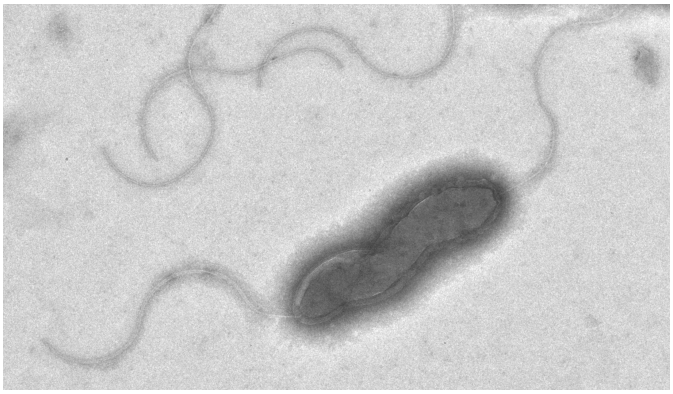


Figure 1.2: Transmission electron micrograph of *C. jejuni* wild type strain NCTC 11168. Taken by Stoakes (2017).

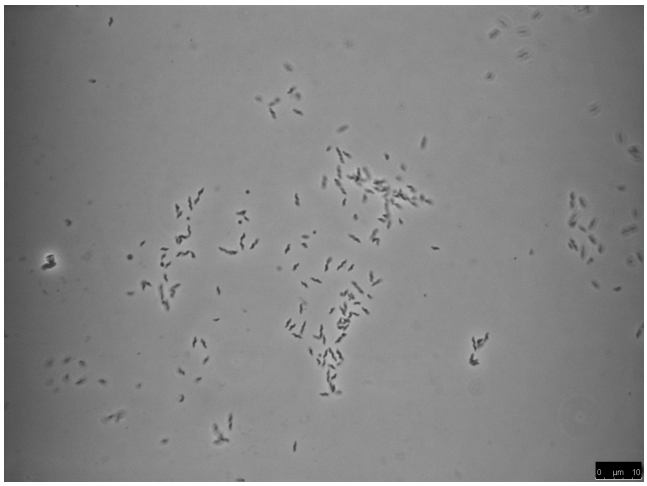


Figure 1.3: Microscopy image of *C. jejuni* wild type strain NCTC 11168. Taken on a Leica microscope at 100X resolution. Scale shown on bottom right.

1.2.1 Epidemiology and transmission of *C. jejuni*

The *Campylobacter* spp. is widespread in the environment and found in many different types of hosts and niches around the world, primarily in avian species including but not limited to domestic livestock, wild animals, surface water, and as an indirect contaminant of fresh produce as reviewed in Kaakoush *et al.* (2015) and Johnson *et al.* (2017). Insects such as the adult house fly and tenebrionid beetle have also been reported as transmission vectors (Taema *et al.*, 2008). Chicken meat is a major source of *C. jejuni*, the predominant species infecting humans (Skarp *et al.*, 2016). Although there are other ways to contract *C. jejuni*, the main route of transmission, particularly in developed countries, is via consumption of undercooked chicken meat as illustrated in studies by Sheppard *et al.* (2009) in Scotland and Workman *et al.* (2005) in Barbados. The latter found that chicken meat was frequently contaminated by *C. jejuni*, and there was a high isolation rate from chicken faecal specimens (Workman *et al.*, 2005).

As reviewed in Kaakoush *et al.* (2015), other ways to contract *C. jejuni* involve handling and/or contacting infected animals, ingesting other contaminated food products such as unpasteurised milk, drinking contaminated water, and less commonly via the faecal-oral route from contact between infected persons. In particular, in developing countries campylobacteriosis (infection with *Campylobacter* spp.) is endemic in children under the age of 5 due to lack of infrastructure and access to clean water and is therefore a significant burden. However, most likely due to immunity, infected children surviving into adulthood have asymptomatic carriage (Kaakoush *et al.*, 2015). Consumption of raw and untreated fresh produce is also a potential source of *Campylobacter* infection most notably from lettuce and spinach (Mohammadpour *et al.*, 2018).

It has been reported that *C. jejuni* is viable at 4 °C even after four months when in the VBNC state (Rollins and Colwell, 1986). This could explain how it is able to persist in water environments and is able to survive refrigeration temperatures important for transmission to humans. Another mechanism that allows *C. jejuni* to survive in harsh environments and thus aiding in transmission is its ability to form biofilms. When under aerobic stress there was an increase in biofilm formation in *C. jejuni* (Reuter *et al.*, 2010). It is important to note that although *C. jejuni* remains viable in such environments, outside of its optimal niches it is not capable of replicating, especially below the temperature of warm-blooded animals at ≈30 °C (Perez-Perez and Kienesberger, 2013).

1.2.2 Sequelae of Campylobacteriosis

Campylobacteriosis is a term to describe infection with any member of the *Campylobacter* genus. Around 90 % of campylobacteriosis cases are due to the species *C. jejuni*, while the remaining 10 % are mainly by *Campylobacter coli* and very rarely other *Campylobacter* species (Bronowski *et al.*, 2014). The typical manifestations of campylobacteriosis are very similar to symptoms of bacterial gastroenteritis caused by other aetiological agents such as *Salmonella*, *Shigella*, and *Yersinia* spp. making it difficult to distinguish (WHO, 1980). Therefore, it is highly likely that many incidences of campylobacteriosis are actually under-reported.

Campylobacteriosis causes gastroenteritis manifesting as abdominal pain, with acute diarrhoea which may be bloody, and can be accompanied with a fever (Hou *et al.*, 2012). In rare cases, complications may arise post-infection. Reactive arthritis can be triggered by gastroenteritis caused by *C. jejuni* (Hannu *et al.*, 2002). Campylobacteriosis is also a major risk factor for the neuropathological disorders Guillain-Barré syndrome (GBS) and its variant Miller-Fischer syndrome (MFS). It is strongly hypothesised that molecular mimicry between *Campylobacter* lipooligosaccharides (LOS) and human gangliosides is responsible for these autoimmune disorders (Ang *et al.*, 2001).

The infectious dose of *Campylobacter* is very low with 500 bacterial cells reported to be sufficient for causing infection and illness (Robinson, 1981). There was a dose-related response observed during human infection with higher doses correlating with higher rates of infection and shorter incubation times (Tribble *et al.*, 2010). This has also been reviewed by Teunis *et al.* (2018) where it is speculated that higher doses are more likely to cause illness and sequelae than lower doses.

1.2.2.1 Current treatments

Usually, campylobacteriosis is self-limiting and there is no need for antibiotic treatment. In some cases though, treatment is required (where fluoroquinolines were the first line drug of choice), and these include but are not limited to: HIV positive patients, pregnancy, immunocompromised patients, and severe and/or prolonged sequelae (Acheson and Allos, 2001). Over the years there has been an increase in antibiotic resistance in *C. jejuni* as reviewed in Nichols *et al.* (2012) and Hou *et al.* (2012). Fluoroquinolone resistance has been on the rise due to its intensive use in livestock and has resulted in *C. jejuni* being listed as a Category 2 priority pathogen for development of new antibiotics by the World Health Organization (WHO) (Sproston *et al.*, 2018; WHO, 2018). *C. jejuni* is intrinsically resistant to many antibiotics and can develop resistance rapidly to other drugs on the market, but so far erythromycin, which has a low resistance rate as well as many other advantages, is currently the best drug of choice (Acheson and Allos, 2001). Alternatives such as bacteriophage and probiotics could also be potential treatments, and development of a vaccine has been contemplated (Johnson *et al.*, 2017; Riddle and Guerry, 2016).

1.2.3 Host colonisation by *C. jejuni*

1.2.3.1 Colonisation in avian species as a commensal

It is widely considered that *C. jejuni* is a natural commensal of chickens with asymptomatic carriage in the gut, in particular the caecum (Pielsticker *et al.*, 2012). There was a pro-inflammatory response generated when colonising chickens but no pathology was observed even with high numbers of bacterial cells present (Smith *et al.*, 2008). However, a study by Humphrey *et al.* (2014) demonstrates a prolonged inflammatory response in broiler chickens from *C. jejuni* infection, causing mild diarrhoea, though this was hypothesised as a result of impaired immunity from poor standards of living. Broiler chickens are especially susceptible to colonisation, with *Campylobacter* spreading through the flock over a few days after the initial chick is infected with up to 10^8 CFU/g bacterial cells found in the caecum (Hermans *et al.*, 2011b). Successful colonisation is dependent on a number of factors which are extensively reviewed in Hermans *et al.* (2011a). These include genes involved in a myriad of stresses such as temperature regulation and heat shock, iron regulation, oxidative and nitrosative stress, as well as virulence factors such as motility. The reasons for asymptomatic carriage in the chicken gut as opposed to human infection are not well understood. Human and chicken isolates vary in colonising the chicken gut suggesting that human isolates may have lost the ability to colonise chickens (Pielsticker *et al.*, 2012).

1.2.3.2 Colonisation in mammalian species as a pathogen

Many mammalian species can be infected with *C. jejuni* as confirmed in the study by Taema *et al.* (2008), which looked at the number of species that were *Campylobacter* positive at a zoo based in the UK between 1990 and 2003. There seemed to be a link with temperature and increased infection with seasonality possibly playing a role in the infections of mammals at the zoo (Taema *et al.*, 2008). Seasonality also occurs with human infection related to the behaviour of certain age groups and the type of transmission and local climate (Strachan *et al.*, 2014).

Long term and chronic persistent infection is possible in humans as demonstrated by Bloomfeld *et al.* (2018), although this is normally associated with an underlying immunocompromised state. A study by Crofts *et al.* (2018) found variants of genes were selected for during human infection and especially in persistent and recurrent infections. These include the 'ON' variant of the phase-variable *cipA* (cell invasion protein A) responsible for modification of the flagella which is a major virulence factor (see section 1.2.3.4), a truncated version of *cmeR* which represses the expression of the multridrug efflux pump *cmeABC* due to two single nucleotide polymorphisms (SNPs) resulting in constitutive expression of the pump, and *pseD* encoding for the protein PseD which attaches pseudaminic acid to flagellin in the 'OFF' state in 1/3 of recurrent cases (Crofts *et al.*, 2018).

1.2.3.3 Intracellular survival and invasion of host cells

C. jejuni can be internalised by non-phagocytic epithelial cells once it reaches the site of infection residing in *Campylobacter*-containing vacuoles (CCVs) for up to 48 hours (Watson and Galán, 2007; Bouwman *et al.*, 2013). Survival of *C. jejuni* inside host cells are dependent on avoiding delivery to lysosomes and adapting to an oxygen-limited environment whilst inside a CCV (Watson and Galán, 2007). As CCVs do not appear to be recognised by host epithelial cells or cause any damage, and intracellular replication is yet to be determined, this may be a way for *C. jejuni* to avoid detection and evade the host innate immune system (Taylor *et al.*, 2015). Entry into the cell was thought to be dependent on caveolae, a type of lipid raft found on the host cell membrane but this was refuted as new evidence indicates that *C. jejuni* binds to fibronectin which leads to activation of the Epidermal Growth Factor (EGF) and focal adhesion kinase (FAK) which is part of the focal complex of the host cell (Konkel *et al.*, 2013).

There have been conflicting reports on which side of polarised epithelial cells is an entry point for *C. jejuni*. Bouwman *et al.* (2013) revealed *C. jejuni* invasion of Caco-2 cells was on the basal cell side independent of host actin and microtubule rearrangement, which are typically seen with other intracellular invasive bacteria, suggesting a unique and novel entry mechanism (Bouwman *et al.*, 2013). However, Taylor *et al.* (2015) only observed CCVs on the apical side and allude this as a consequence of differences between *in vitro* and *in vivo* cell invasion studies (Taylor *et al.*, 2015). Motility is thought to be important for entry into host cells since a mutagenesis screen that resulted in non-motile insertional mutants were non-invasive and non-adherent, although there were also fully motile mutants that were also defective in adherence and invasion, indicating external factors such as adhesins that play a role in motility-dependent invasion (Yao *et al.*, 1994). It is ambiguous whether motility or flagella per se are responsible for host cell internalisation but factors described in the following sections illustrate why flagella may be crucial for invasion. Many other factors have been implicated for invasion including capsule production, lipooligosaccharide sialylation, and chaperone proteins (Cróinín *et al.*, 2012).

1.2.3.4 Motility and flagella

C. jejuni have amphitrichous flagella (one at either end of the cell) which supports corkscrew like motion for invasive infection (Inoue *et al.*, 2018). Compared to other flagellated bacteria, the basal body contains many more stator complexes allowing motility through highly viscous fluids (Radomska *et al.*, 2016). In addition to being a major virulence factor, motility is also important for bacterial survival: to reach their niche site of residence, avoid harsh conditions, and to search for nutrients (Roure *et al.*, 2012). As well as motility and invasion, the flagella also acts as a transport system for secreting virulence proteins termed *Campylobacter* invasion antigens (Cia) (see section 1.2.3.5) as *C. jejuni* lack the typical type III secretion system (T3SS) employed by other Gram-negative enteropathogens (Guerry, 2007). Mutations that lead to attenuated motility render *C. jejuni* unable to invade host cells (Golden and Acheson, 2002).

Flagella are assembled in a hierarchical manner via, early, mid, and late genes and is therefore growth-phase dependent. It was shown that the three sigma (σ) factors present in *C. jejuni* regulate flagellar biogenesis in the following order: σ^{70} mainly involved in early housekeeping genes that are constitutively expressed forming the base, σ^{54} encodes the mid genes for the rod and hook, and σ^{28} encodes for late gene products that assemble the filament (Carrillo *et al.*, 2004). The early genes: FliF which forms the MS ring, the rotor component of the C ring FliG, and the GTPase FlhF are required for the activation of the two-component system FlgRS, which can increase the activity of σ^{54} RNA polymerase holoenzyme to express σ^{54}-dependent genes (Boll and Hendrixson, 2013). FlgM is an anti-σ^{28} repressor that is normally secreted after completion of the hook to allow for expression of σ^{28}-dependent late genes (Boll and Hendrixson, 2013). However, this was not the case in *C. jejuni*; FlgM controls the length of the filament by binding to σ^{28} (also known as FliA) in a temperature-dependent manner and suppresses both σ^{54} and σ^{28}-dependent genes (Wösten *et al.*, 2010b). Disruption of FlgM, which is a small middle gene product of $\approx$7 kDa, did not show a clear phenotype at 37 °C but at 42 °C there was increased transcription of the filament protein FlaA resulting in longer than normal flagella and reduced motility, although there was no difference in the transcription of FlgM between the temperatures (Wösten *et al.*, 2010a). FlgM was also detected in the supernatant of the wild type bacteria at 42 °C so it was proposed that σ^{28}/FliA is perhaps more active at higher temperatures and changes conformation and thereby dissociates from FlgM (Wösten *et al.*, 2010a). The length of *C. jejuni* flagella may be attributed to FlaG, as a deletion mutant of FlaG resulted in longer than usual flagella, and it is thought that FlaG negatively regulates flagellar assembly by suppressing σ^{28} (Inoue *et al.*, 2018).

Regulation of flagellar assembly is tightly controlled as it is a metabolically expensive process (El Abbar *et al.*, 2019). Though mainly mediated by the alternative sigma factors σ^{54} and σ^{28}, and the FlgRS two-component system, other factors have been shown to be involved in flagellar assembly. The global Carbon storage regulator/Regulator of secondary metabolism (CsrA) was found to interact with the flagellar assembly factor FliW in *C. jejuni* (Radomska *et al.*, 2016; Dugar *et al.*, 2016). CsrA mainly targets and represses the messenger RNA (mRNA) of the major flagellin *flaA*, and is antagonised by FliW (Dugar *et al.*, 2016). CsrA is an RNA-binding protein and was shown to target the 5'UTR of *flaA* in *C. jejuni* (Fields *et al.*, 2016). Specific amino acids in CsrA binding site were found to be crucial for binding to the 5' UTR of *flaA* (El Abbar *et al.*, 2019). Once the flagella is assembled, post-translational modifications contribute to the function and activity of the flagella. For instance, a deletion mutant of the Per-Ant-Sim (PAS) domain containing Cj1387c which has now been renamed as HeuR by Johnson *et al.* (2017) was found to have an increase in autoagglutination and it is speculated that this was attributed to differences in glycosylation (Reuter *et al.*, 2015).

Post-translational glycosylation occurs across all domains of life and the N- and O-linked glycosylation systems present in *C. jejuni* are similar to their counterparts in eukaryotes (Szymanski *et al.*, 2003). *C. jejuni* and *C. coli* flagellins are the most heavily glycosylated proteins in bacteria, although this varies from strain to strain (Guerry *et al.*, 2006; Ewing *et al.*, 2009). There are three flagellin genes in *C. jejuni*: *flaA* regulated by σ^{28}, *flaB* regulated by σ^{54}, and *flaC* which is a secreted non-structural protein regulated by σ^{70} that is thought to play a role in host cell invasion, with FlaA and FlaB assembled as the major components of the flagellar filament (Parkhill *et al.*, 2000; Radomska *et al.*, 2016; Inoue *et al.*, 2018). *C. jejuni* flagella are O-linked glycosylated with pseduaminic acid and its derivatives, whereas N-linked glycosylation commonly occurs on membrane proteins (Hitchen *et al.*, 2010; Alemka *et al.*, 2013).

Site-specific insertional mutants of *pseB*, a pseudaminic acid biosynthesis gene resulted in non-glycosylated flagellins that could not be assembled and were non-motile, thus glycosylation is essential for flagellar biogenesis (Goon *et al.*, 2003). Furthermore, a mutant of the *pdxA* gene responsible for production of pyridoxal-5'-phosphate, better known as vitamin B6, also impairs pseudaminic acid production and hence glycosylation of flagella, as vitamin B6 possibly acts as a co-factor in pseudaminic acid production (Asakura *et al.*, 2013). The variation in flagellar glycoproteins is due to the phase-variable gene Cj1295 containing a homopolymeric-tract (Hitchen *et al.*, 2010). A glycosylation island that is present in live-stock associated strains, including the reference strain NCTC 11168 but not non live-stock associated strains such as 81-176, was found to decorate flagellin with legionaminic acid, which are implicated in autoagglutination and hydrophobicity of flagella (Howard *et al.*, 2009). Deletion mutants of the entire glycosylation island resulted in lower levels of colonisation in the chicken gut suggesting that legionaminic acid decoration of flagella is a major factor for chicken colonisation and commensalism (Howard *et al.*, 2009). The function of CipA was previously unknown and is normally in the 'OFF' variant but was found to be turned 'ON' during human recurrent infections only and revealed to modify flagella (Crofts *et al.*, 2018). Therefore, it seems that post-translational modification of the flagella is a major factor to whether *C. jejuni* acts as a commensal or pathogen.

1.2.3.5 Cia proteins

Co-cultivation of *C. jejuni* with intestinal cells leads to the production and secretion of a series of virulence proteins called Cia proteins (Rivera-Amill *et al.*, 2001). It is not certain what the exact function(s) of Cia proteins are but it has been reported that they are involved in invasion of host cells (Cróinín *et al.*, 2012). The first identified Cia protein was CiaB, which was found to have similar properties to T3SS proteins and is essential for uptake of *C. jejuni* strain F38011 cells into INT407 host cell lines and for the secretion of all Cia proteins (Konkel *et al.*, 1999). CiaB lacks a signal sequence and contact with the host cell is required for the secretion of Cia proteins, which are typical traits of the T3SS (Konkel *et al.*, 2004). The only T3SS found in the *C. jejuni* genome was the flagellar apparatus, which has been confirmed to be responsible for the secretion of Cia proteins (Christensen *et al.*, 2009; Neal-McKinney and Konkel, 2012). However, a mutant of CiaB in strain 81-176 did not have an effect on invasion of INT407 cells, implying the roles

of Cia proteins vary between strains (Guerry, 2007). A number of genes encoding for Cia proteins were up-regulated under sodium deoxycholate stress including CiaB in strain F38011 (Malik-Kale *et al.*, 2008). One of these genes encodes for a protein that was highly secreted in a reporter strain and identified as CiaC and found to be important for maximal host cell internalisation (Malik-Kale *et al.*, 2008; Christensen *et al.*, 2009). Though the exact process of CiaC delivery is yet to be determined, a functional flagella and host-cell contact is needed for delivery of CiaC into the host cytosol in strain F38011, however, not all Cia proteins have the same fate as some that were exported via the flagella were not delivered to the cytosol (Neal-McKinney and Konkel, 2012).

1.2.4 Metabolism of *C. jejuni*

Metabolism plays an important role in survival and pathogenicity as strains of *C. jejuni* that are better at colonisation encode for more genes related to catabolising substrates (Hofreuter, 2014). Factors that target host processes to increase nutrient availability or to promote nutrient uptake are considered as nutritional virulence factors (Gao *et al.*, 2017). There also has to be some degree of metabolic flexibility in order for *C. jejuni* to adapt to all the different hosts and environments it encounters (Stahl *et al.*, 2012). *C. jejuni* was thought to be incapable of metabolising carbohydrate sugars as demonstrated by the lack of genes to facilitate degradation of these energy sources (Parkhill *et al.*, 2000). These include the absence of glucokinase and phosphofructokinase of the Embden-Meyerhof-Parnas pathway and missing enzymes in Entner-Doudoroff pathway and oxidative pentose phoshate pathway (Vorwerk *et al.*, 2014; Hofreuter, 2014; Vegge *et al.*, 2016; Gao *et al.*, 2017). However, a small subset of strains including *C. jejuni* subsp. *doylei* do have the full set of genes (Vegge *et al.*, 2016).

A logical alternative carbon source to sugars are amino acids, although only a limited number are catabolised in *C. jejuni* (Guccione *et al.*, 2008). There is an order of preference depending on the availability of amino acids: serine, aspartate, asparagine, glutamate, and proline which is only used as a last resort (Stahl *et al.*, 2012). Indeed, insertions in the *sdaC* gene disabling serine uptake had severe growth defects in rich media but mutations in other amino acid uptake systems did not hinder growth (Gao *et al.*, 2017). These amino acids, with the exception of asparagine, are commonly found in chicken faecal matter (Stahl *et al.*, 2012). *C. jejuni* is also able to utilise intermediaries from the tricarboxylic acid (TCA) cycle (also known as the citric acid cycle or Krebs cycle) such as pyruvate as a primary carbon source (Stahl *et al.*, 2012; Vorwerk *et al.*, 2014).

Recently it was discovered that L-fucose, a common sugar found in the intestine, is uptaken and metabolised by some strains of *C. jejuni* giving it an advantage over other commensal bacteria, though this pathway is not activated during colonisation of the chicken caeca (Stahl *et al.*, 2011). L-fucose is a constituent of mucin, which is a chemoattractant for *C. jejuni* (Hermans *et al.*, 2011a). The strain NCTC 11168 utilises the fucose pathway and a study by van der Hooft *et al.* (2018) assessed its metabolic response to the addition of glutamic acid, commonly found in processed food, and fucose to this strain. NCTC 11168 was cultured in minimal essential medium alpha modification (MEMα) and it was found that serine, proline, aspartate,

glutamate, and glutamine were depleted as expected but also phenylalanine, which has not been observed before (van der Hooft *et al.*, 2018). The addition of glutamic acid or fucose drastically changed the number of metabolites produced in particular with fucose, which was predominated by sulfur-containing metabolites (van der Hooft *et al.*, 2018). A study by Rasmussen *et al.* (2013) also suggested that peptides are also another alternative carbon source.

1.3 Genetic features and regulatory systems in *C. jejuni*

1.3.1 National Collection of Type Cultures 11168 and its variants

The National Collection of Type Cultures (NCTC) 11168 was originally isolated from a patient suffering from diarrhoea in 1977 and is the first *C. jejuni* strain to be sequenced in 2000 (Parkhill *et al.*, 2000; Skirrow, 1977). The data revealed many features in the tightly packed A/T rich 1,641,481 base pair (bp) circular genome. There is a lack of repeated sequences and only three predicted sigma factors compared to seven in *Escherichia coli* (Parkhill *et al.*, 2000; Chaudhuri *et al.*, 2011).

Due to numerous passaging in the laboratory environment, the sequenced strain showed signs of deficiency colonising one-day-old chicks and severe combined immune deficiency (SCID) mice, and was found to have impaired motility (Gaynor *et al.*, 2004). Fortunately, the original stock of the clinical isolate (strain 5636/77) was frozen without any known sub-culturing and comparatively colonises one-day-old chicks very well and is highly motile (Gaynor *et al.*, 2004). The strain used in this project is similar to the original clinical isolate and has seven point mutations all non-synonymous apart from the mutation in Cj0455c when compared to the non-motile sequenced reference strain by Parkhill *et al.* (2000) (Table 1.1) (Stoakes, 2017). There were no mutations present in intergenic regions and all mutations apart from *cipA* and *hisG* have actually been previously identified in other studies that have sequenced their original clinical isolate strain of NCTC 11168, so these mutations can be considered as SNPs (Jerome *et al.*, 2011; Cooper *et al.*, 2013; Butcher and Stintzi, 2013; Pascoe *et al.*, 2019). When compared to the original clinical isolate and other sequenced strains, our strain has two novel mutations in *cipA* and *hisG* (Cooper *et al.*, 2013).

Gene	Position in reference	Reference nucleotide	Nucleotide change	Amino acid change
mreB	253191	A	G	D → G
cheA	262345	A	G	I → T
Cj0431	393542	T	A	* → K
Cj0455c	420550	A	G	-
cipA	638796	T	G	V → K
Cj0807	760188	A	G	K → E
hisG	1525525	G	A	G → D

Table 1.1: Mutations in our NCTC 11168 laboratory strain compared to the NCBI non-motile reference genome. * indicates a stop codon and - indicates there was no change. Adapted from (Stoakes, 2017).

1.3.2 Gene expression and transcriptional regulation

The central dogma of biology is that DNA is transcribed into RNA, which is then translated into protein and the latter is irreversible as seen in Figure 1.4 (Anonymous, 1970). This process is present in all forms of life and is fundamental for cellular functions. DNA is the template of life storing all the biological information needed to make the molecular machinery essential for maintenance and growth of an organism. A gene is a sequence of nucleotides (either DNA or RNA), which can code for a functional molecule or protein determining the order of the nucleotide sequence that consequently determines the amino acid primary structure of proteins. A gene is expressed when it is transcribed and/or further translated to a protein or other functional gene product. In bacteria, transcription occurs simultaneously with translation. But not all genes are expressed at once; expression can be regulated at each level of the central dogma, especially at the level of transcription, and genes are essentially turned 'ON' or 'OFF' or fined-tuned depending on the stage of development or the environment the cells are in. In fact, some genes may never be expressed but are still present in the chromosome, for example, if they are recessive or from leftover remnants of evolution.

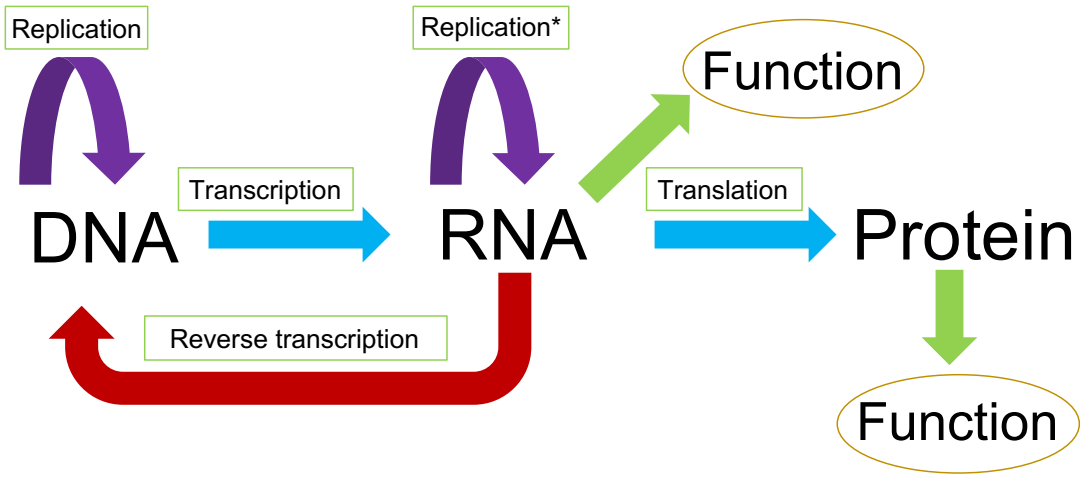


Figure 1.4: Schematic diagram of the central dogma of biology. A simplified model describing the events of genetic processing. DNA is transcribed to RNA, which can either be functional or translated to a functional protein. RNA can be reverse transcribed to DNA and replicated though the latter* only occurs in RNA viruses. DNA can also be replicated.

The stages of transcription in both prokaryotes and eukaryotes are as follows: initiation, escape, elongation, and termination. Initiation is the stage which is the most tightly regulated (Browning and Busby, 2016) Promoters are regions of DNA sequence in the genome where RNA polymerase binds to, to synthesise DNA to RNA and are typically upstream of genes. While normally regulated by transcription factors and sigma factors present in the RNA polymerase complex, they have been found within open reading frames (orf) and in other non-coding regions of DNA (Wade and Grainger, 2018). This gives rise to spurious or pervasive transcription. Promoters are generally A/T rich since there are only two hydrogen bonds, which makes it easier for RNA polymerase to separate the two DNA strands to allow for transcription as opposed to three hydrogen bonds with C and G nucleotides.

1.3.2.1 *C. jejuni* sigma factors

The sigma (σ) subunit of the RNA polymerase holoenzyme complex determines which promoters the complex binds to for transcription initiation to occur. As mentioned previously, there are only three predicted sigma factors in NCTC 11168: sigma factor σ^{70} (RpoD) the housekeeping protein responsible for the majority of transcription, sigma factor σ^{54} (RpoN) and sigma factor σ^{28} (RpoF/FliA) which are both mainly involved in flagellar biosynthesis (Chaudhuri *et al.*, 2011). The three stages of flagellar assembly are controlled by the sigma factors with the early genes regulated by σ^{70}, middle genes by σ^{54}, and late genes by σ^{28} (Le *et al.*, 2015). In bacteria, the housekeeping σ^{70} sigma factor normally binds to the -10 (bp upstream) box (also known as the pribnow box) and the -35 (bp upstream) box, but in genome-reduced (i.e. small genome) bacteria, the promoters have gradually lost the -35 box element (Miravet-Verde *et al.*, 2017). This is also the case in *C. jejuni* where there is a strong periodic variation of A/T stretches upstream of the pribnow box rather than the -35 box element (Petersen *et al.*, 2003). It is speculated that the loss of the -35 box in genome-reduced bacteria could be due to the small number of sigma factors raising the question of what the determining factor for transcription initiation and recognition by RNA polymerase is (Miravet-Verde *et al.*, 2017).

Gene expression regulation via sigma factors is linked to resilience to stress. An RpoN σ^{54} mutant in the strain 81-176 made in Hwang *et al.* (2011a)'s study showed that the mutant had defects in motility and was more susceptible to acid and osmotic stresses, but had no difference between the wild type with temperature or antimicrobial resistance, and strangely was more resistant to oxidative stress. The transcriptional profile of an RpoN σ^{54} mutant in the strain NCTC 11168 had unsurprisingly many flagellar genes that were down-regulated but also the gene Cj1242, which was previously annotated as having an unknown function that was later identified as the invasion antigen CiaC (Chaudhuri *et al.*, 2011). A number of intergenic regions were also expressed with the potential to be non-coding RNAs (ncRNAs), suggesting a possible role in the regulation of σ^{54}-dependent genes (Chaudhuri *et al.*, 2011).

1.3.2.2 Spurious transcription

Spurious or pervasive transcription derived from spurious promoters result in a low-level background of transcription as reviewed in Wade and Grainger (2018). As the genomes of bacteria are mostly protein-coding, novel transcripts are mainly identified antisense to genes or within genes, but as there is no determined regulatory function unless further validation experiments are employed, it is unclear whether these transcripts can be considered pervasive or *bona fide* (Lybecker *et al.*, 2014). In some cases spurious transcripts can be beneficial but more likely than not detrimental to the cell (Wade and Grainger, 2014). There are many mechanisms to counteract this including minimising promoter-like sequences within genes through natural selection so that the A/T content in genes is reduced (Wade and Grainger, 2018). However, *C. jejuni* has a very A/T rich genome (Parkhill *et al.*, 2000). Not all spurious transcripts are bad; over time these can evolve into functional RNAs such as small RNAs (sRNAs) (section 1.3.3) (Wade and Grainger, 2018). Inefficient termination can also lead to readthrough which is another source of spurious transcription

(Wade and Grainger, 2014).

1.3.2.3 RNA turnover

RNA degradation and turnover in bacteria are important and necessary for rapid adaptation to environmental conditions, so that transcripts which are no longer needed can be degraded leaving resources for translation available for gene products that enable survival (Hör *et al.*, 2018). *C. jejuni* does not encode for RNase E or RNase II which are responsible for majority of the RNA degradation in many Gram-negative bacteria, but does encode for polynucleotide phosphorylase (PNPase) which degrades mRNA in the 3'-5' direction (Haddad *et al.*, 2012). A mutant of the gene *pnp* encoding for PNPase was shown to have a protective role against cold temperatures, but further to this in a separate study a Δ*pnp* also had motility defects (Haddad *et al.*, 2009, 2012). Ribonuclease III (RNase III) is another RNA degradation enzyme that is highly conserved in prokaryotes and eukaryotes and requires a divalent metal ion preferentially Magnesium (Mg^{2+}) as a co-factor (Haddad *et al.*, 2013). RNase III is active in a range of temperatures and pH levels in *C. jejuni* and may play a role in regulating PNPase levels (Haddad *et al.*, 2013). RNase R was also found to be active in a wide range of pH levels, temperatures, and low salt concentrations with a preference for KCl but is also able to cleave DNA (Haddad *et al.*, 2014). During nutrient depleted conditions such as transition to stationary phase or starvation, stable RNAs such as ribosomal RNA (rRNA) are degraded due to a decrease in growth possibly to release nutrients from ribosomes (Deutscher, 2003).

1.3.3 Small RNAs (sRNAs) in *C. jejuni*

sRNAs are found in all forms of life and although they were first discovered in 1981 in *E. coli*, research in bacterial sRNAs have only recently come to fruition with the advent of next-generation sequencing (Stougaard *et al.*, 1981). Generally non-coding, sRNAs modulate gene expression by directly binding to mRNA transcripts to repress translation or in some cases enhance translation as seen in Figure 1.5. Cis-acting sRNAs are encoded antisense to the target gene and have perfect base-pairing with their target, whereas trans-acting sRNAs are encoded elsewhere and are normally intergenic and only have partial complementarity to their target.

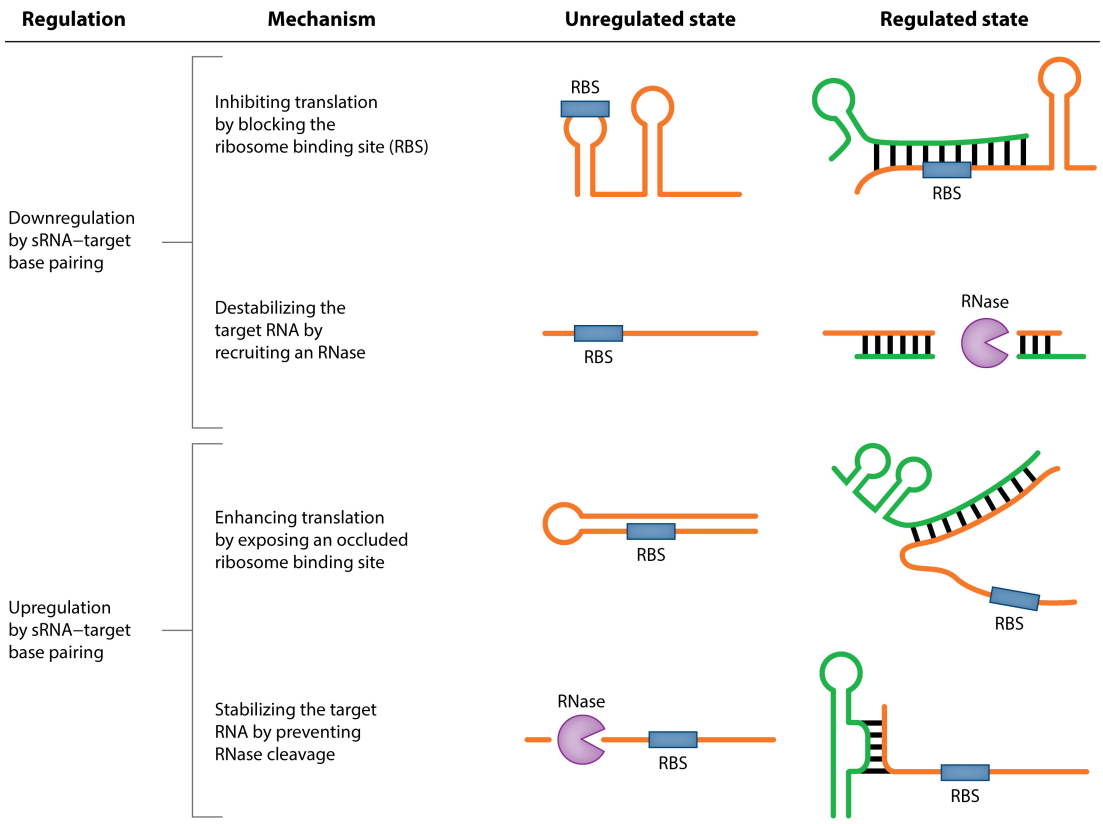


Figure 1.5: Schematic diagram of trans-acting sRNAs interacting with their targets. The different ways sRNAs are able to block or enhance translation. Taken from Nitzan *et al.* (2017)

In *C. jejuni*, sRNAs are not characterised very well. *C. jejuni* does not encode or possess a homologue of the sRNA chaperone Hfq protein which is present in many Gram-negative bacteria (Chao and Vogel, 2010). It does have a homologue of another RNA-binding protein CsrA, however, in *C. jejuni* CsrA acts independently by binding directly to mRNA rather than interacting together with sRNAs (Fields *et al.*, 2016; Dugar *et al.*, 2016). Since *C. jejuni* lack many of the typical regulatory factors that other bacteria possess, sRNAs could play a major role in post-transcriptional regulation. Since sRNAs are mainly non-coding with a few exceptions, they can also be referred to as ncRNAs and these terms are used interchangeably in the literature. It is important to note though that ncRNA is a very broad and general term and that transfer RNA (tRNA) and ribsomal RNA (rRNA) can also be considered ncRNAs. For this project, ncRNA refers to RNA transcripts found which have the potential to be sRNAs.

Several putative ncRNA candidates have been identified in *C. jejuni* by primary transcriptome studies and RNA-sequencing (RNA-seq) data, though only a few have been extensively studied. The role of two σ^{28} ncRNAs: CjNC1 and CjNC4 identified by our data and both published primary transcriptome studies in Dugar *et al.* (2013) and Porcelli *et al.* (2013) and named according to the terminology from the latter study, were characterised by Le *et al.* (2015). Both paralogues were predicted to target 5'UTR of σ^{54} genes and are conserved throughout many *C. jejuni* species, although no clear phenotype was determined from overexpression and deletion of these ncRNAs (Le *et al.*, 2015).

1.3.4 Phase variation

Phase variation is an alternative regulatory system bacteria employ to alter gene expression rapidly. Insertion or deletion of repeat units in the poly G/C tracts due to slip-strand mispairing is the main mechanism of phase variation in *C. jejuni* resulting in an 'ON'/'OFF' state due to frameshifts (Bayliss *et al.*, 2012; Anjum *et al.*, 2016; Crofts *et al.*, 2018). Therefore, in a population of cells, phase variation can result in different phenotypes more so than the normal mutation rate, increasing diversity and leading to increased resistance against stresses and the host immune system. *C. jejuni* has 29 poly G/C tracts that have seven or more repeats with three residing in intergenic regions, two in pseudogenes, and one at the end of a reading frame (Lango-Scholey *et al.*, 2016). Phase-variable genes seem to play a role in flagellar biogenesis. The FlgRS two-component system which controls the expression of σ^{54}-dependent flagellar genes was found to be to phase-variable and specific to *C. jejuni* with a loss of single nucleotide leading to a minority of the population to have an 'OFF' FlgR variant resulting in aflagellate cells (Hendrixson, 2006). This was proposed as a mechanism for spreading to new hosts as aflagellated variants are more prone to shedding in faecal matter due to impaired colonisation (Hendrixson, 2006). The flagellin glycosylation locus in NCTC 11168 contains 10 genes that have homopolymeric tracts with one gene Cj1295 characterised as contributing to variable flagellin glycan structures which result in differing mobilities, possibly to help evade the host immune system (Hitchen *et al.*, 2010). The phase-variable gene *cipA*, which was found to have a role in flagellar modification is normally in the 'OFF' variant and was found to be turned 'ON' in relapse human infections (Crofts *et al.*, 2018). Furthermore, nine other phase-variable genes were also conserved in sequenced strains isolated from relapse patients (Crofts *et al.*, 2018).

1.3.5 Two-component systems

In bacteria, gene expression can be modulated by two-component systems consisting of a sensor kinase triggered by an environmental stimuli which then subsequently activates the response regulator (Mitrophanov and Groisman, 2008). Relative to other bacteria, *C. jejuni* have comparably fewer two-component systems; five of the seven cognate two-component systems predicted from Parkhill *et al.* (2000)'s study have been characterised (Luethy *et al.*, 2015): PhosRS activated by phosphate limitation (Wösten *et al.*, 2006), FlgRS activated by the flagellar export apparatus (Joslin and Hendrixson, 2009), DccRS activated by late stationary growth phase (Wösten *et al.*, 2010b), CprRS involved in biofilm formation (Svensson *et al.*, 2015), RacRS important for avian colonisation, heat shock response, motility, and cell length (Brás *et al.*, 1999; Apel *et al.*, 2012). Another unnamed two-component system, CJJ81176_1484 (Cj1492c) and CJJ81176_1483 (Cj1491c) in the strain 81176, has also recently been characterised and shown to repress many genes involved in metabolic processes (Luethy *et al.*, 2015). Finally, *C. jejuni* contains a CheAY two-component system that is part of the chemotaxis system triggered by the activation of methyl-accepting chemotaxis proteins, which detect chemoeffector signals that attract bacteria (Zautner *et al.*, 2012).

1.3.6 Global regulators and responses to stress in *C. jejuni*

C. jejuni is exposed to many environments and niches as it is transmitted from one animal host to another. There are also microenvironments within each host that *C. jejuni* traverses before reaching the site of colonisation/infection. Despite encountering many hostile conditions in the environment during transmission, due to its limited genome size *C. jejuni* lack the typical coping mechanisms and many homologues of global regulators that other food-borne pathogens possess, most notably RpoS which is a global stress regulator present in many Gram-negative bacteria (Navarro Llorens *et al.*, 2010; Parkhill *et al.*, 2000). Many genes and transcriptional regulators involved with adaptation to stress are missing compared to the model organisms *E. coli* and *Bacillus subtilis* for oxidative stress, osmoregulation, stationary phase, heat and cold shock, quorum sensing, and global regulation as listed and reviewed in Park (2002).

Nevertheless, there are still many global regulators in *C. jejuni* that are necessary for regulation, especially in response to stresses, and may have bigger or alternate roles than their counterparts in other species. One major global regulator is the ferric uptake regulator (Fur), which primarily regulates iron acquisition genes and exists in the apo-form but can bind with ferrous iron (Fe^{2+}) and conform to the holo-form, and was shown to also regulate energy metabolism, formation of the cell membrane, oxidative stress, and acid stress (Askoura *et al.*, 2016). The regulatory mechanisms OxyR and SoxRS regulating oxidative stress are highly conserved in Gram-negative bacteria but are not found in *C. jejuni*, instead PerR (peroxide resistance regulator) which is a homologue to the metalloregulator Fur is present in *C. jejuni*, the first known Gram-negative to do so (Kim *et al.*, 2015).

Although implicated as playing a role for stationary phase survival, CsrA was actually found to regulate virulence associated proteins (Fields *et al.*, 2016). As mentioned previously, CsrA is heavily involved in the regulation of flagella but also regulates the oxidative stress response, biofilm formation, and host cell invasion (Fields and Thompson, 2008). A *csrA* mutant had increased expression of the global transcriptional regulators Fur, CosR, and RacR, although there was not much overlap between the regulons (Fields *et al.*, 2016). This implies there is a lot of crosstalk between regulatory networks that are very complex. There is also a possibility that there are more transcriptional regulators yet to be discovered in *C. jejuni*. Recently, two novel regulators were identified in the *C. jejuni* strain 81176 named CheP and CheQ, which regulate the chemotactic *cheVAW* operon and are homologues to Cj0248 and Cj0249 in NCTC 11168 respectively, and were previously annotated as having an unknown function (Cha *et al.*, 2019). A number of growth and stress conditions that *C. jejuni* encounters in its life and the regulatory mechanisms that governs them will be discussed in the next sections.

1.3.6.1 The stringent response

When bacteria undergo stress due to limited nutrient availability, the stringent response is normally activated to ensure their survival and persistence (Gaca *et al.*, 2015). A deficiency in nutrients including but not limited to: amino acid starvation, iron limitation, entry into stationary phase, carbon limitation as well as other cues such as temperature change, variation in oxygen levels, and acid stress appear to trigger the stringent response (Atkinson *et al.*, 2011; Ronneau and Hallez, 2019). Two alarmone signalling molecules are involved in the stringent response - guanosine 5'-diphosphate 3'-diphosphate (ppGpp) and guanosine 5'-triphosphate 3'-diphosphate (pppGpp) collectively known as (p)ppGpp (Boutte and Crosson, 2008). An accumulation of (p)ppGpp leads to a number of genes required for growth to be repressed and redirects resources to express genes involved in resistance and survival, and nutrient acquisition to alleviate the stress (Gaca *et al.*, 2015). These alarmone molecules are regulated by the RSH (RelA-SpoT homologue) superfamily enzymes that either synthesise (p)ppGpp from ATP, GTP, or GDP; or degrade back to GTP/GDP and pyrophosphate (PPi) (Irving and Corrigan, 2018). In *E. coli* where the stringent response was first discovered, the RSH enzymes RelA synthesises (p)ppGpp whereas SpoT is bifunctional and is able to both synthesise and degrade (p)ppGpp (Atkinson *et al.*, 2011). Gram-negative bacteria typically have both RelA and SpoT, whereas in Gram-positive bacteria there is only a single RelA/SpoT homologue that is bifunctional (Irving and Corrigan, 2018). *C. jejuni* only has one of the two RSH superfamily enzymes SpoT (Cj1272c) which is thought to have dual-functionality due to the absence of a RelA homologue akin to Gram-positive bacteria, and the deletion mutant Δ*spoT* in strain 81-176 showed there was a defect in adherence, invasion, and intracellular survival (Gaynor *et al.*, 2005).

There are a few identified proteins that interact with SpoT and are linked to the stringent response. For instance the acyl carrier protein (ACP) is post-translationally modified to the holo-ACP activated form and acts as a co-factor for fatty acid synthesis and lipid metabolism and interacts with SpoT (Battesti and Bouveret, 2006). However, this was only found in the Gram-negative *E. coli* and *Pseudomonas aeruginosa* but not in Gram-positive *Streptococcus pneumoniae* and *B. subtilis* (Battesti and Bouveret, 2009). This suggests that the ACP-SpoT interaction may only be present in bacteria that have two RSH superfamily enzymes e.g. possessing RelA and SpoT, although some bacteria such as *C. jejuni* are Gram-negative but only have one RSH superfamily enzyme SpoT. Another protein that was found to interact with SpoT is CgtA (also annotated as Obg in other bacteria), a conserved GTPase that interacts with the 50S ribosomal subunit in various bacteria, and depletion of CgtA in *Vibrio cholerae* demonstrated increased levels of ppGpp (Raskin *et al.*, 2007). *C. jejuni* encodes for the phase-varibale CgtA, which also encodes for the gene product NeuA1, a CMP-NeuAC synthetase (Guerry *et al.*, 2002; Day *et al.*, 2017). It is not clear whether SpoT in *C. jejuni* interacts with these proteins and further studies are needed to clarify this.

Inorganic polyphosphates (poly-P) are found ubiquitously in living cells formed of long chains of phosphate residues, which are covalently linked by the high energy releasing phosphoanhydride bonds that are also found in ATP (Candon *et al.*, 2007; Achbergerová and Nahálka, 2011). Thus, they serve as energy sources and a supply of phosphates for many cellular functions, and accumulate inside granules (Kumar *et al.*, 2016). In addition, poly-P was shown to be involved in tolerance and adaptation of various stresses and is linked to the stringent response, as low levels of (p)ppGpp resulted in lower levels of poly-P in *E. coli* stringent response mutants (Kuroda, 2006; Candon *et al.*, 2007; Gangaiah *et al.*, 2009). The synthesis and degradation of poly-P is mediated by the highly conserved polyphosphate kinase 1 (PPK1) with many bacteria also containing a second polyphosphate kinase 2 (PPK2) (Kumar *et al.*, 2016). The reaction of both polyphosphate kinases are reversible, though PPK1 favours the synthesis of poly-P from ATP whereas PPK2 preferentially produces GTP from poly-P (Malde *et al.*, 2014). *C. jejuni* contains homologues of both kinases with the deletion mutant of PPK1 Δ*ppk1* in strain 81176 showing decreased formation of VBNC cells in acid compared to the wild type (Gangaiah *et al.*, 2009). In *E. coli*, poly-P can form a complex with the ATP-dependent Lon protease to break down cytoplasmic proteins in order to release amino acids for the synthesis of stress-response proteins (Achbergerová and Nahálka, 2011). Two other proteases ClpAP and ClpXP also degrade proteins in *E. coli*, and the mutations of ClpAP, ClpXP, and Lon have the same phenotype as PPK mutants (Kuroda, 2006). It was also found that the poly-P-Lon complex specifically degrades certain ribosomal proteins which are not normally degraded by Lon in the absence of poly-P (Kuroda, 2006). *C. jejuni* has homologues of Lon and Clp proteases named as ClpP and ClpX; both Lon and ClpP are necessary for optimal growth at high temperatures and invasion of epithelial cells, although the link between ATP proteases and polyphosphate kinases have not been investigated in *C. jejuni* (Cohn *et al.*, 2007).

Exopolyphosphatases (PPX) are involved with the homeostasis of poly-P as they degrade poly-P into smaller chains of inorganic phosphates (Pi), but some have a dual function and also act as a guanosine pentaphosphate phosphohydrolase (GPPA) which generates and contributes to the pool of ppGpp (Malde *et al.*, 2014). In *E. coli*, (p)ppGpp was found to inhibit PPX which in turn leads to an accumulation of poly-P and subsequent binding to Lon (Kuroda, 2006). *C. jejuni* contains two PPX/GPPA enzymes and deletion mutants of these enzymes affects: survival in osmotic stress, motility and biofilm formation, and invasion and intracellular survival, and *spoT* and *ppk1* were up-regulated in the mutants possibly compensating for PPX/GPPA's role in maintaining poly-P and ppGpp levels (Malde *et al.*, 2014). However, extremely high levels of poly-P are toxic to the cell as was shown in Rudat *et al.* (2018), where several mutated alleles of *ppk* in *E. coli* accumulated large amounts of poly-P in rich media and had dramatic decrease in growth compared to the wild type and also had even higher levels of poly-P after nutrient limitation. This was due to poly-P sequestering metal ions, in particular magnesium, which is limited in Luria broth (LB) used to grow the strains (Rudat *et al.*, 2018).

1.3.6.2 Stationary phase and nutrient deprived conditions

There have been seven described distinct growth stages of bacteria by Buchanan (1918), although the five most commonly referred to are lag phase, logarithmic (log)/exponential phase, stationary phase, death phase, and long-term stationary phase (Rolfe *et al.*, 2012). Lag phase refers to the initial stage where there is latent period of growth, exponential phase when bacterial cells are rapidly dividing, stationary phase when bacteria stop replicating though there may still be metabolic activity, and death phase when there is a decline in bacteria cell numbers. In the natural environment, bacteria are subjected to nutrient-limited conditions as well as other stresses which trigger entry into stationary phase, although in the laboratory this is mainly due to depletion of nutrients in the media (Jaishankar and Srivastava, 2017). During mid to late stationary phase, *C. jejuni* exhibits an acetate switch where there is uptake and utilisation of acetate rather than excretion of this by product due to the depletion of acetate-producing carbon sources such as L-serine or pyruvate (Wright *et al.*, 2009).

Bacterial resistance to stress is normally increased as exponential phase transitions to stationary phase, but this is not the case in *C. jejuni* (Turonova *et al.*, 2017). Many Gram-negative bacteria confer higher resistance to bile salt stress at early stationary phase but in contrast, *C. jejuni* does not fare well under bile salt stress at stationary phase when compared with exponential phase (Kelly *et al.*, 2001). Unlike most Enterobactericeae, *C. jejuni* lacks the RpoS sigma factor also known as σ^B or σ^{34}, which is essential for entry into the stationary phase for bacteria such as *E. coli* (Kelly *et al.*, 2001). As mentioned in the previous section, survival during stationary phase was in fact derived from the stringent response regulated by *spoT* for *C. jejuni* (Gaynor *et al.*, 2005). In late stationary phase, the two-component system DccRS proteins are also activated (Wösten *et al.*, 2010b).

During transmission and in some microenvironments in the host, *C. jejuni* have to tolerate nutrient deprived conditions such as starvation. Nutrient deprivation was shown to significantly affect culturability and viability the most as well as reduced invasion of epithelial cells when compared with heat shock and oxygen exposure (Mihaljevic *et al.*, 2007). Cj0917 is a homologue to CstA (carbon starvation protein A) found in *E. coli* and is up-regulated in late stationary starvation-like conditions relative to exponential phase in *C. jejuni* (Rasmussen *et al.*, 2013). A CstA mutant has multiple phenotypes including reduced motility (Rasmussen *et al.*, 2013).

1.3.6.3 Temperature

C. jejuni is a mesophile and the optimal temperature for *C. jejuni* growth is 37 °C to 42 °C which is the core body temperature of humans and avian species respectively, as well as covering a range of core body temperatures for other mammals (Haddad *et al.*, 2009). A study by Stintzi (2002) found that during the 5 °C shift from 37 °C to 42 °C many genes were differentially expressed, hinting that different hosts and intrinsic temperatures can affect the behaviour of *C. jejuni*. Chemotaxis and quorum-sensing was shown to be elevated during growth at 37 °C compared to 42 °C suggesting that the human body

temperature is better suited for virulence traits (Khanna *et al.*, 2006). Outer membrane vesicles (OMVs) are one alternative method for *C. jejuni* to secrete virulence proteins as it lacks a T3SS employed by other Gram-negative enteropathogens, and it was found that genes encoding for OMVs were differentially regulated between 42 °C and 37 °C in *C. jejuni* strain 81176 (Taheri *et al.*, 2019). Many virulence-associated proteins were enriched at 37 °C with the lipoprotein FlgP found exclusively at 37 °C and three proteins: CJJ81176_pTet0044 Cpp47, CJJ81176_1366 (Cj1364c) Fumarase C, and CJJ81176_1418 (Cj1419c) a SAM-dependent methyltransferase found exclusively at 42 °C (Taheri *et al.*, 2019).

As mentioned previously, *C. jejuni* is viable at 4 °C. However, it cannot replicate or grow at low temperatures due to the absence of cold shock proteins (Hazeleger *et al.*, 1998). PNPase seems to play a role in survival at 4 °C and it has been implied that clinical isolates are more resistant to low temperatures than poultry isolates (Haddad *et al.*, 2009). Normally, when exposed to low temperatures, the fatty acid composition of the cell membrane is altered to cope with the temperature change as seen in *E. coli*, but this in turn increases its sensitivity to heat as the membrane has to be reorganised to cope with any sudden subsequent changes to high temperatures (Hughes *et al.*, 2009). As this does not occur in *C. jejuni*, there is no necessity to reorganise the fatty acid composition of the membrane when encountering high temperatures making *C. jejuni* slightly more tolerant to high temperatures after chilling (Hughes *et al.*, 2009). This attribute could explain how *C. jejuni* manages to be successful as a food-borne pathogen.

During food processing of broiler chickens, *C. jejuni* present on the carcass can undergo the scalding process at high temperatures ranging from 50-64 °C (Zhuang *et al.*, 2013). A study by Nguyen *et al.* (2006) showed a poultry isolate L51 had a maximum rate of cell death at 57 °C, and at 55 °C there was a decrease in log D-value (decimal reduction time) of 4.6 minutes. In response to elevated temperatures, bacteria induce a heat shock response resulting in many up-regulated heat-shock proteins that consist of chaperones or proteases to help protect denaturation of proteins (Maleki *et al.*, 2016). The reference genome NCTC 11168 contains many heat-shock proteins including GroEL, GroES, GrpE, DnaK, DnaJ, and several ATP-proteases (Andersen *et al.*, 2005; Parkhill *et al.*, 2000). Furthermore, *lon*, *groESL*, *clpB*, and *dnaK* were also up-regulated in response to heat stress (Svensson *et al.*, 2008). Heat shock proteins are regulated by the repressors HrcA and HspR which have homologues in *C. jejuni*, however, the alternative sigma factor σ^{32} in model organisms such as *E. coli* which adds an additional layer of heat shock regulation is not present in *C. jejuni* (Holmes *et al.*, 2010). The high temperature requirement A (HtrA) protein was found to be essential for *C. jejuni* growth at 44 °C and above, as it degrades and prevents protein aggregates but is also necessary for oxygen resistance and invasion and migration through host epithelial cells (Boehm *et al.*, 2015). Deletion mutants of the two-component system RacRS also showed colony forming growth defects at 44 °C although this could likely be due to failure of inducing the heat shock response and possible mis-regulation of *dnaJ*, which RacRS normally represses (Apel *et al.*, 2012).

1.3.6.4 Oxidative and anaerobic environments

Regarded as an obligate microaerophilate, *C. jejuni* prefers low oxygen levels but must cope with oxidative stress, especially whilst in the environment between hosts where there can be a fluctuation in the levels of oxygen. Upon exposure to oxygen, many reactive oxygen species (ROS) are formed, for example: superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($HO^\bullet$) which damage nucleotides and other molecules in bacterial cells (Atack and Kelly, 2009; Hwang *et al.*, 2011b; Gundogdu *et al.*, 2016). These ROS are formed from the incomplete reduction of oxygen to water:

$$O_2 \rightarrow O_2^- \rightarrow H_2O_2 \rightarrow HO^\bullet \rightarrow H_2O$$

Bacteria can be subjected to endogenous ROS produced within the cell and exogenous ROS from host cells (Imlay, 2019). There is a basal level of scavenging enzymes such as the KatA catalase which can keep endogenous ROS in check, whereas a large influx of radicals will trigger the activity of global regulators (Imlay, 2019). Other major detoxification and scavenging proteins in response to ROS have also been identified as stated in Flint *et al.* (2014): AhpC - alkyl hydroxperoxide reducatase, SodB - superoxide dismutase, Tpx - thiol peroxidase, Bcp - thiol peroxidase, Dps - bacterioferritin, MsrA/B, and Cj1386 - ankyrin-containing protein (Flint *et al.*, 2014).

The regulation of oxidative stress in *C. jejuni* is very convoluted and involves several regulators which have overlapping regulons. A reduction in CosR (Cj0355c) (*Campylobacter* oxidative stress regulator) levels increased the resistance of *C. jejuni* to oxidative stress, whereas an overexpression of the regulator decreased its resistance (Hwang *et al.*, 2011b). CsrA was also found to regulate genes essential for survival in oxidative stress (Fields and Thompson, 2008). The regulators of response to peroxide RrpA (Cj1546) and RrpB (Cj1556) which were previously unknown genes in the reference genome NCTC 11168 have now been re-annotated as part of the MarR family of transcriptional regulators (Gundogdu *et al.*, 2015). Although many *C. jejuni* strains do contain RrpA, only a limited number of strains such as NCTC 11168 have RrpB as well as RrpA, and most of these strains tend to be associated with livestock and are part of the MLST clonal complex type ST-21 and type ST-61 (Gundogdu *et al.*, 2016). A deletion mutant of RrpB was found to be sensitive to oxidative stress with the complemented strain reviving oxidative resistance (Gundogdu *et al.*, 2011). Unusually, *C. jejuni* strains that only have RrpA tend to have higher resistance to oxidative stress and antimicrobial resistance (Ugarte-Ruiz *et al.*, 2018). The two-component system CprRS, which stands for *Campylobacter* planktonic growth regulation, was also implemented in oxidative stress with the deletion mutant Δ*cprS* having up-regulation of many oxidative stress genes (Svensson *et al.*, 2009). To maintain copper homeostasis, *C. jejuni* encodes two genes: *copA* a multicopper oxidase and *cueO* a P_{1B-1}-type ATPase transporter and copper (I) oxidase (Gardner and Olson, 2018). Mutants of both *copA* and *cueO* had susceptibility to oxidative stress and it was speculated that toxic levels of copper may be reacting with ROS (Gardner and Olson, 2018).

C. jejuni encounters anaerobic conditions in the host intestine and in particular in the chicken caecum which was found to be populated by obligate anaerobes (Weingarten *et al.*, 2008). To avoid these unfavourable conditions, *C. jejuni* preferentially colonises sites with higher oxygen tension such as the mucous layer, intestinal crypt, or the region between the mid small intestine and mid colon (Hofreuter, 2014). However, *C. jejuni* does have the potential to grow under anaerobic respiration if nitrite reductase, which acts as an alternative electron acceptor and nitrate are not compromised and this seems to play an important role in chicken colonisation (Weingarten *et al.*, 2008). In fact, many alternative electron donors such as fumarate, trimethylamine-*N*-oxide (TMAO), dimethyl sulfoxide (DMSO), as well as nitrite and nitrate are able to support anaerobic growth (Sellars *et al.*, 2002).

1.3.6.5 Nitrosative stress

Nitric oxide (NO) is a free radical employed by macrophages as a defence against bacterial infection, with its ability to freely diffuse across cell membranes to form reactive nitrogen species that interfere with many bacterial cellular functions, react with thiols, and inhibit iron-sulfur molecules (Flatley *et al.*, 2005). Levels of NO are elevated during infection with *C. jejuni* demonstrating the bactericidal activity of NO (Elvers *et al.*, 2005). The single domain globin Cgb helps to scavenge and detoxify NO from *C. jejuni* and was found to be regulated by the Nitrosative stress sensing Regulator (NssR) (Elvers *et al.*, 2005). NssR also mediates the expression of a truncated globin Ctb which scavenge NO, though it is not as potent as Cgb (Avila-ramirez *et al.*, 2013). When *C. jejuni* bacterial cells were cultured under oxygen-limited conditions similar to anaerobic conditions, it was found that NssR and Cgb were not up-regulated and did not provide any protection against NO itself, and that there is a possibility of another reactive nitrogen species generated from NO that is not susceptible to globins (Avila-ramirez *et al.*, 2013).

1.3.6.6 Acid shock

C. jejuni inevitably encounters a low pH environment as it passes through the gastrointestinal tract through gastric acid of hosts to reach the site of infection and intracellularly within phagosomes inside host cells (Varsaki *et al.*, 2015). To counteract this, bacteria employ strategies to: prevent entry of H^+, pump and export out H^+, absorb H^+, or mitigate damage from acid shock, though these have not been characterised in *C. jejuni* (Birk *et al.*, 2012; Reid *et al.*, 2008). The *C. jejuni* strain CI 120, a poultry processing isolate, had an adaptive tolerance response to acid stress, although this was not seen in the reference strain NCTC 11168 (Varsaki *et al.*, 2015). In NCTC 11168, the global metalloregulator Fur, known to regulate iron homeostasis, was found to play a role in survival during acid stress as the deletion Δ*fur* mutant had lower cell viability compared to the wild type (Askoura *et al.*, 2016). Furthermore, there is cross-protection against oxidative stress upon acid exposure mediated by Fur and the up-regulation of the catalase gene *katA* (Askoura *et al.*, 2016). Although unclear, it was hypothesised that flagella may play a role in resistance to acid stress by allowing *C. jejuni* to rapidly move away to more ideal conditions as many flagellar biogenesis genes were up-regulated upon exposure to acid (Askoura *et al.*, 2016; Reid *et al.*, 2008). This

includes moving towards a protective host such as amoebae since acidic conditions in the surrounding environment induces migration towards and uptake of *C. jejuni* by *Acanthamoeba polyphaga* which can act as a transmission vector (Axelsson-Olsson *et al.*, 2010). A study by Le *et al.* (2012) also found that *C. jejuni* could survive up to 30 minutes at pH 3.5 and there was an induction of σ^{54} induced motility genes. Furthermore, it is speculated that the drop in pH can also be an indicator that the bacteria are inside a host thereby triggering the expression of virulence determinants such as flagella (Reid *et al.*, 2008). *C. jejuni* cells primed to acid stress also increased survivability in the secondary stresses starvation and osmotic stress and enhanced adherence and invasion of intestinal epithelial cells (Kumar-Phillips *et al.*, 2013).

1.3.6.7 Bile

Bile is found in high concentrations in mammalian intestines secreted by the liver and plays a role in digestion as well as having an antimicrobial effect, particularly against Gram-positive bacteria (Cremers *et al.*, 2014). Bile is produced from cholesterol and mainly consists of bile salts/acids, small amounts of phospholipids, and cholesterol (Raphael *et al.*, 2005; Svensson *et al.*, 2008; Urdaneta and Casadesús, 2017). As well as the host, commensal bacteria also produce bile acids namely lithocholic acid and deoxycholic acid (Negretti *et al.*, 2017). Bile is able to traverse the outer membrane of Gram-negative bacteria or pass through porins, interfering with cell membrane integrity and causing further damage once inside the cell, damaging DNA and proteins due its detergent nature (Raphael *et al.*, 2005; Urdaneta and Casadesús, 2017). The two common bile salts are cholate and deoxycholate derivatives. Despite its antibacterial properties, bile salts are useful as an environmental signal for bacteria to perceive the intestinal environment (Urdaneta and Casadesús, 2017). Although considered as a chemorepellent, Li *et al.* (2014) found that this was dependent on the origin of bile, for example human, murine, or bovine bile were actually chemoattractants for *C. jejuni* (Hugdahl *et al.*, 1988; Li *et al.*, 2014). Sodium deoxycholate was found to induce synthesis of Cia (described in section 1.2.3.5), specifically the *ciaB* promoter (Malik-Kale *et al.*, 2008). A study by Negretti *et al.* (2017) also discovered that deoxcholate triggers an oxidative-like stress response as the levels of ROS were increased 12.2 fold in the presence of 0.05 % deoxycholate. This is likely due to the alteration of the membrane permeability of cells affecting the electron transport chain and contributing to the production of ROS (Negretti *et al.*, 2017).

OMVs were mentioned previously to be differentially regulated between different host temperatures and having different protein contents. Under a low concentration of bile, the production of OMVs or viability of cells was not affected but influenced the protein content of OMVs, with some proteins exclusive to the stress (Taheri *et al.*, 2018). To deal with bile stress, the operon *cmeABC* encoding for the multidrug efflux pump (CME), which confers resistance to antibiotics and heavy metals, is also able to pump out bile salts and is essential for *C. jejuni* colonisation in chickens (Hermans *et al.*, 2011a). A mutant of the *Campylobacter* bile resistance regulator CbrR, which is a two-component system response regulator, rendered strain F38011 sensitive to 0.05 % sodium deoxycholate and was also revealed to be important for bile resistance (Raphael *et al.*, 2005).

1.3.6.8 Hyperosmotic stress

The osmolarity of intestines in the host changes as *C. jejuni* passes through depending on the area and/or food ingested. Salt, generally in the form of NaCl, is commonly used as a food preservative and antibacterial agent causing hypertonic osmotic stress as *C. jejuni* is transmitted via food processing (Burgess *et al.*, 2016). The increase in osmolarity in the environment from the addition of salt and/or other solutes disrupts the integrity of the cell membrane and causes dehydration due to an efflux of water caused by osmosis, and conversely the decrease in osmolarity results in an influx of water that leads to swelling and possible lysis (Wood, 2015; Burgess *et al.*, 2016). Bacteria can also encounter dehydration and dessication in the environment during transmission. A study by Cameron *et al.* (2012) looked at the sensitivity of *C. jejuni* strain 81-176 to different solutes and concentrations. The study found that 2.0 % NaCl was lethal to *C. jejuni* and rather than triggering the VBNC state, hyperosmotic stress induced filamentation of cells at 1.0 and 1.5 % NaCl, although at 2.0 % cells did not display filamentation and were non-motile (Cameron *et al.*, 2012). In addition, the deletion of the sigma factor RpoN (σ^{54}) mutant had increased sensitivity to osmotic stress exhibiting growth defects (Hwang *et al.*, 2011a). Moreover, poly-P has been implicated as vital for survival in osmotic stress as a deletion mutant of the polyphosphate kinase Δ*ppk1* which synthesises poly-P in 81-176, had significant decrease in growth compared to wild type under media supplemented with 0.25 M NaCl (Candon *et al.*, 2007).

1.3.6.9 Iron acquisition

Iron is used by all forms of life, and is essential for bacterial survival. It is the most abundant metal inside the human body although the availability is withheld to serve as a broad innate defence against pathogens termed as nutritional immunity, as well as its low solubility and toxicity to the host (Hood and Skaar, 2012). Iron is part of the iron-sulfur complex in the active sites of many key enzymes involved in a myriad of metabolic networks in *C. jejuni*, and is therefore crucial for growth and survival (Stahl *et al.*, 2012). Bacteria require on average 10^{-7}-10^{-5} M of iron for survival, and inside a host *C. jejuni* has to compete with the gut microbiota as well as the host for iron acquisition (Kortman *et al.*, 2014). This metal is not freely available in the host but exists intracellularly in haemoglobin or myoglobin cells (Sheldon *et al.*, 2015). The levels of iron can fluctuate due to diffusion from food or from inflammatory responses from damaged cells (van Vliet *et al.*, 2008). However, high concentrations of iron can be toxic for bacteria due to the production of radical species and ROS so it is important to have a precisely controlled regulatory system in place for the uptake of iron (Butcher *et al.*, 2012). Iron induced ROS are generated via the Fenton and Haber-Weiss reactions:

1. **Fenton reaction:**

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + HO^{\bullet} + OH^{-}$$

2. **Iron reduction:**

$$Fe^{3+} + O_2^{-} \rightarrow Fe^{2+} + O_2$$

3. **Haber-Weiss reaction:**

$$O_2^{-} + H_2O_2 \rightarrow O_2 + HO^{\bullet} + OH^{-}$$

with products from the first two reactions as inputs for the third reaction (Kortman *et al.*, 2014). Therefore, too much or too little iron is detrimental to the cell. Iron-limited conditions could also be an indication that *C. jejuni* is inside a host and used as a signal to express virulence factors (Holmes *et al.*, 2005).

Fur is present in most Gram-negative bacteria including *C. jejuni* and regulates iron acquisition genes in response to the levels of intracellular iron (Butcher *et al.*, 2015). PerR, which is a Fur homolog mentioned previously that regulates oxidative stress, is also involved in iron regulation, and single and double knock-out mutants that were made demonstrated the crosstalk between the two metalloregulators (Butcher *et al.*, 2015). There is no surprise that the Fur and PerR regulons overlap with each other as ROS including hydroxides (OH^{-}) are formed by the Fenton and Haber-Weiss reactions involving iron and oxygen (Butcher *et al.*, 2015). So as well as the precise regulation of iron levels, genes involving detoxifying peroxides and radicals need to correspond to the presence of iron to limit the damage caused from the by-products of the Fenton and Haber-Weiss reactions (Kortman *et al.*, 2014; Butcher *et al.*, 2015).

C. jejuni contains many uptake systems that are able to acquire iron from a variety of sources. The most common states of iron are in the ferrous (Fe^{2+}) and ferric (Fe^{3+}) forms and due to the oxygen levels in the gut it is likely that ferric iron may be reduced to ferrous iron as seen in the above reaction, although the levels of ferrous iron in the gut are unknown (Naikare *et al.*, 2006; Krewulak and Vogel, 2008). Ferrous iron is able to freely diffuse through the outer membrane of Gram-negative bacteria and is one of the important sources of iron, as disruption of the Fe^{2+} transporter in the cytoplasmic membrane of various gut pathogens were found to have attenuated colonisation in the gut (Naikare *et al.*, 2006). *C. jejuni* encodes for a FeoAB Fe^{2+} transporter that is repressed by iron and is important for intracellular survival (Naikare *et al.*, 2006). Bacteria can secrete siderophores which are small Fe^{3+} chelators and uptake the iron-bound siderophore (Behnsen and Raffatellu, 2016). To date, *C. jejuni* does not seem to produce its own siderophores yet has an uptake system to pirate siderophores from other gut bacteria (Raines *et al.*, 2016).

The transport system to uptake siderophores through the outer membrane in Gram-negative bacteria is mediated by the TonB-ExbB-ExbD complex that provides proton motive force energy to the ligand on the outer membrane and spans the periplasm (van Vliet *et al.*, 2008; Naikare *et al.*, 2013). *C. jejuni* encodes for three TonB-ExbB-ExbD systems that are iron-regulated and two outer membrane transporters CfrA and CfrB that depend on TonB-ExbB-ExbD with the former appearing to have broad recognition for many siderophores (Palyada *et al.*, 2004; Zeng *et al.*, 2013; Naikare *et al.*, 2013). CfrA is able to

uptake enterobactin which is a siderophore produced by *E. coli* and once inside the periplasm, it was thought that a periplasmic binding protein CeuE binds to enterobactin to allow uptake through the cytoplasmic membrane (Raines *et al.*, 2016). However, a study by Raines *et al.* (2016) discovered that the presence of another protein named Cee in the periplasm of *C. jejuni* hydrolyses enterobactin, and that CeuE actually has higher affinity for the hydrolysis product $[Fe(bisDHBS)]^{2-}$, which gives *C. jejuni* a competitive advantage as enterobactin is also prone to hydrolysis in aqueous solution (Raines *et al.*, 2016).

Inside the host, free iron can be sequestered intracellularly to prevent uptake from pathogens and to retain iron availability for the host to use, causing an iron-limited extracellular environment (Richard *et al.*, 2019). Ferric iron can be sequestered in haem in the form of haemin and other haem-associating proteins, and since bacteria are only capable of uptaking haem, the proteins have to be separated from the haem groups (Ridley *et al.*, 2006). The rupture of red blood cells which contain haemoglobin (haemolysis) is a source of haem and this can be achieved from toxins released from pathogenic bacteria to induce haemolysis (Krewulak and Vogel, 2008; Richard *et al.*, 2019). *C. jejuni* is capable of obtaining haem from haemin, haemoglobin, haemin-haemopexin, and haemobglobin-haptoglobin (Pickett *et al.*, 1992; Miller *et al.*, 2009). This is mediated by the Fur-regulated operon *chuABCDZ* where *chuA* is the outer membrane receptor, *chuBCD* encodes for an ATP-binding cassette (ABC) transport system on the cytoplasmic inner membrane, and *chuZ* which is transcribed divergently from the other genes is a haem oxygenase that degrades haem to release iron (Ridley *et al.*, 2006; Johnson *et al.*, 2016). Normally, the inner membrane receptor is TonB-ExbB-ExbD dependent, though this has not been verified for the *C. jejuni chuBCD* transport sytem (Ridley *et al.*, 2006; Richard *et al.*, 2019). *chuA* was revealed to be vital for haem uptake but the ABC transport system *chuBCD* was not, implying redundancy of transport systems across the inner membrane (Ridley *et al.*, 2006). A regulator protein containing a PAS domain named as HerR in the *C. jejuni* strain 81176 was also found to be essential for haem uptake and both HeuR and ChuA is needed as loss of one or the other resulted in no haem uptake (Johnson *et al.*, 2016). The homologue of HeuR in the reference genome NCTC 11168 Cj1387c was found to affect flagella-flagella interactions but many additional genes were differentially expressed in a *heuR* mutant strain in 81176 including oxidative resistance as well as iron, indicating a broader role for HerR (Reuter *et al.*, 2015; Johnson *et al.*, 2016).

Extracellular iron can be complexed with high affinity to the transferrin family of glycoproteins, for example, transferrin that can be found in serum or lactoferrin in mucosal secretions with only variations in glycosylation between the proteins in the host (Baker *et al.*, 2002). Unlike siderophores and haem which can be taken into the bacterial periplasm as whole molecules, iron has to be extracted from transferrin or lactoferrin prior to uptake (Krewulak and Vogel, 2008). Transferrin-bound iron was demonstrated to be taken up when in close proximity by the outer membrane receptor Cj0178 designated as *Campylobacter* transferrin-bound iron utilisation receptor (CtuA) (Miller *et al.*, 2008). CtuA was shown to be essential for chick and rabbit colonisation (Miller *et al.*, 2009). Cj0177 is in an operon upstream of CtuA but it is not clear whether it is involved in transferrin/lactoferrin uptake or haem uptake (Stahl *et al.*, 2012). Associated and encoded divergently to Cj0177 and CtuA is the *cfbpABC* operon, which stands for the ferric binding

protein and altogether encodes for an ABC transporter (Tom-Yew *et al.*, 2005). Unusually, *cfbpABC* can bind iron without a synergistic anion unlike its counterparts in other bacterial species and has a preference for ferrous iron over ferric iron so is therefore established as a unique and new class of ferric binding proteins (Tom-Yew *et al.*, 2005).

Two other iron uptake systems found in *C. jejuni* are not very well characterised. *p19*, Cj1658, and Cj1660-1663 seem to be involved in the uptake of ferric-rhodotorulic acid, a siderophore produced by fungus (Miller *et al.*, 2009). Uptake of ferric-rhodotorulic acid into the periplasm is from an unknown outer membrane receptor, and the siderophore possibly interacts with the periplasmic binding protein *p19* before transportion through the cytoplasmic membrane via Cj1661-1663, which resembles an ABC transporter (Miller *et al.*, 2009; Stahl *et al.*, 2012). However, it is uncertain whether *C. jejuni* would encounter such siderophores in the environment during transmission and it is possible that this uptake system is for a more relevant but unknown siderophore that is similar to rhodoturulic acid (Miller *et al.*, 2009). In strain 81-176, it was found that the homologues to *p19*, Cj1658, and Cj1660-1663 were involved in response to acid, streptomycin, and oxidative stress and these genes were highly up-regulated during human colonisation suggesting additional regulation by other factors alongside iron and a more important position for this uptake system (Liu *et al.*, 2018). Finally, in certain strains of *C. jejuni*, Cj0444 seems to play a role in iron uptake as strains with Cj0444 as a pseudogene (e.g. NCTC 11168) contain functional CfrA and Cj0178 outer membrane receptors, but strains where Cj0444 is functional do not contain CfrA and Cj0178, suggesting a substitutional role for Cj0444 (Miller *et al.*, 2009).

1.4 Rationale for the project

1.4.1 Transcriptional landscape of NCTC 11168

To date, in literature there have been many studies investigating the transcriptional landscape of *C. jejuni* including strain NCTC 11168. These include primary transcriptome studies by Dugar *et al.* (2013) and Porcelli *et al.* (2013). A number of RNA-seq and microarray studies under specific conditions have also been carried out including but not limited to: iron limitation and repletion conditions, hyperosmotic stress, colonisation of a natural host, and an RpoN mutant (Butcher and Stintzi, 2013; Cameron *et al.*, 2012; Taveirne *et al.*, 2013; Chaudhuri *et al.*, 2011). However, the conditions used in these studies were either routine standard growth conditions or only focused on a specific stress, and to date there has not been a comparative study across many different growth and stress conditions. The primary transcriptome studies have revealed many novel features of the architecture of the *C. jejuni* genome. Leaderless mRNAs i.e. mRNA transcripts without 5' UTRs were thought to be rare in bacteria but were found in NCTC 11168 (Porcelli *et al.*, 2013). They do not require conventional ribosomal recruitment for translation to occur (Kaberdina *et al.*, 2009). In particular, all studies which used RNA-seq serendipitously discovered many putative ncRNAs which have the potential to be sRNAs, though many may have been missed as sRNAs are generally induced by stress responses as reviewed in Hoe *et al.* (2013).

1.4.2 Hypothesis and project outline

As there are many missing promoters that are yet to be identified for many genes in the annotated genome and *C. jejuni* is exposed to various stresses during transmission and inside a host, we hypothesise that many promoters could actually be condition-dependent. To the best of our knowledge, the primary transcriptome has only been investigated under standard laboratory conditions at one growth phase. We aim to highlight many additional promoters that were not identified due to their condition-dependent nature. To identify these missing promoters, a compendium of 22 growth and stress conditions that were described in the previous sections were selected that are relevant to colonisation in the chicken host as a commensal, infection in the human host, or transmission between the hosts. The compendium was pooled together for determining the primary transcriptome and we predict that many novel transcripts such as putative ncRNAs will also be identified. Many stress conditions have cross-talk and overlapping regulons, thus the individual RNA samples were also sequenced for comparative analysis of each transcriptional profile. Unusually, *C. jejuni* lacks many hallmark global regulators and established stress-response genes found in other bacteria so there may be another system at play (Svensson *et al.*, 2008). A number of transcriptional profile studies have been described under various conditions but due to different factors such as media, temperature, and strains it is hard to do a comparative analysis between the published studies. Since the compendium includes both existing and novel conditions it is ideal for confirming the data as a reliable resource with potential novel findings.

1.5 Aims and objectives

The overall aim of this research project was to determine the transcriptional landscape of NCTC 11168 under relevant chicken colonisation, human infection, and transmission conditions. This will enable us to find novel transcripts and potentially novel ncRNAs (with the potential to be an sRNA), uncovering clues to the regulatory systems at play. Understanding how environmental factors affecting the molecular mechanisms involved in survival and virulence of *C. jejuni* and how this affects regulation will greatly support advances in other research areas to help combat this pathogen (Turonova *et al.*, 2017).

The objectives of this study were:

1. To optimise RNA extraction methods and select the best suited to identify low-level transcripts such as ncRNAs in NCTC 11168.
2. To characterise TSS using Cappable-seq and determine novel transcripts including ncRNAs from a pool of 22 host-relevant transmission stress conditions.
3. To determine the individual transcriptional response of NCTC 11168 including newly identified ncRNAs from the same compendium of 22 host-relevant transmission stress conditions.

Chapter 2

Materials and methods

2.1 Maintenance and growth of *C. jejuni*

The strain NCTC 11168 from the National Collection of Type Cultures was kindly donated by Professor Charles Penn from the University of Birmingham and was used throughout this project. This version of the strain is from the original clinical isolate and is motile, in contrast to the genome sequenced strain in Parkhill *et al.* (2000)'s study. NCTC 11168 was stored in Microbank bacterial preservation tubes (Fischer Scientific) at -80 °C and routinely cultured on *Campylobacter* blood-free selective agar supplemented with CCDA (containing Cefoperazone and Amphotericin B) (Oxoid) for 24 hours. A loop of NCTC 11168 culture was restreaked onto Mueller-Hinton (MH) (Oxoid) or cation-adjusted Mueller-Hinton 2 (MH2) (Sigma-Aldrich) 1.5 % agar (Bacto^{TM}Agar) plates overnight. All incubation and growth steps were carried out at 37 °C in a variable atmosphere incubator (VAIN) (Whitley VA500 workstation cabinet) (90 % N_2 (v/v), 6 % CO_2 (v/v), 4 % O_2 (v/v)).

2.1.1 Standard growth curves

From MH2 plates, a loopful of NCTC 11168 culture was inoculated into 5 mL of MH2 broth in 25 cm^3 Vented Capped Tissue culture flasks (Falcon). The flasks were incubated in the VAIN shaking at 200 rpm on an orbital shaker (Ika, VXR basic Vibrax®) for 15 hours to 15 hours 30 minutes. The optical density at 600 nm (OD_{600}) was measured on a spectrophotometer (Biochrom) and the calculated volume of liquid culture to achieve an initial OD_{600} of 0.05 was centrifugated at 5000 x *g* for 3 minutes, and resuspended in 4.5 mL of fresh MH2 broth in a 6-well plate (Greiner). 1 mL from each sample well was used to measure the initial OD_{600} resulting in a final volume of 3.5 mL in each well of the 6-well plate (Greiner). Four biological replicates and two blanks were used. The 6-well plate (Greiner) was placed in a FLUOStar Omega microplate reader (BMG labtech) shaking at 200 rpm for 48 hours measuring at intervals of 30 minutes, with the atmosphere control unit (ACU) set to 6 % CO_2, and 4 % O_2.

2.1.2 Growth and stress conditions

Table 2.1 shows the summarised list of all conditions used in this project. Transient stresses were applied after growing *C. jejuni* to exponential phase at 37 °C apart from nutrient starvation, which was applied after reaching early stationary phase. All growth curves were carried out as previously described in Section 2.1.1 but with the appropriate temperature and cells collected at the appropriate OD_{600} detailed in Section 3.2. Unless specified, Mueller Hinton broth 2 (MH2) (Sigma-Aldrich) media was used.

Stress/growth condition	Description of stress	Reference
Body temperature of humans	Grow to exponential phase at 37 °C Grow to early stationary phase at 37 °C Grow to late stationary phase at 37 °C	
Body temperature of chickens	Grow to exponential phase at 42 °C Grow to early stationary phase at 42 °C Grow to late stationary phase at 42 °C	
Iron limitation	Grow in MEM supplemented with 10 M pyruvate, exponential and early stationary phase	(Butcher and Stintzi, 2013)
Iron repletion	Grow in MEM supplemented with 10 M pyruvate and 40 M FeSO, exponential and early stationary phase	(Butcher and Stintzi, 2013)
Bile secretion	Grow in MH2 broth supplemented with 0.1 % sodium deoxycholate, exponential and early stationary phase	(Malik-Kale *et al.*, 2008)
Peroxide stress	Grow to exponential phase in MH2 broth at 37 °C, add 3 mM hydrogen peroxide for 10 minutes	(Klančnik *et al.*, 2006)
Anaerobic stress	Grow to exponential phase in MH2 broth at 37 °C, incubate in anaerobic chamber for 1 hour	
Heat shock	Grow to exponential phase in MH2 broth at 37 °C, incubate at 55 °C for 3 minutes	(Klančnik *et al.*, 2014)
Nutrient starvation	Grow to early stationary phase in MH2 broth at 37 °C, resuspend in Ringer's solution for 5 hours	(Klančnik *et al.*, 2009)
Acid shock	Grow to exponential phase in MH2 broth at 37 °C, resuspend in MH2 broth pH 3.5 for 10 minutes	(Le *et al.*, 2012)
Hyperosmotic stress	Grow to exponential phase in MH2 broth at 37 °C, add 1.5 % NaCl and incubate for 2 hours	(Cameron *et al.*, 2012)
Chicken juice	Grow to exponential phase in MH2 broth at 37 °C and resuspend in 100 % chicken juice for 24 hours at 4 °C Grow to exponential phase in MH2 broth at 37 °C and resuspend in MH2 broth supplemented with 5 % chicken juice for 24 hours at 4 °C	(Birk *et al.*, 2004)
Cold stress	Grow to exponential phase in MH2 broth at 37 °C, resuspend in MH2 broth for 24 hours at 4 °C	(Brown *et al.*, 2014)

Table 2.1 continued from previous page

Stress/growth condition	Description of stress	Reference
Nitrosative stress	Grow to exponential phase in MH2 broth at 37 °C, add 1.5 mM GSNO and incubate for 2 hours	(Elvers *et al.*, 2005)

Table 2.1: List of growth and stress conditions used in this project. References to literature from which the protocol was adapted from are included.

2.1.2.1 Growth of NCTC 11168 with and without supplements

NCTC 11168 growth was optimised in the plate reader (see Section 3.2) and RNA was harvested at exponential, early stationary, and late stationary phases at OD_{600} specified in Section 3.2.2. For 42 °C the temperature of the plate reader was adjusted accordingly and growth curves were characterised. MH2 was supplemented with 0.1 % sodium deoxycholate for growth in bile conditions. Iron-limited conditions were carried out by growing NCTC 11168 in MEMα (Gibco) supplemented with 10 μM pyruvate, and repleted by adding 40 μM $FeSO_4$. Both bile and iron-related conditions were harvested at exponential and early stationary phases.

2.1.2.2 Hyperosmotic and nitrosative stress

NCTC 11168 was grown to OD_{600} 0.45-0.675 (path-length corrected) lower than the exponential harvest point (see Section 3.2.2) to allow for extra growth during incubation. For hyperosmotic stress, cultures were resuspended in 1.5 % NaCl MH2 broth which were filter sterilised with 0.22 μM pore membrane filters (Millipore), whilst S-Nitrosoglutathione (GSNO) (CALBIOCHEM) was added at a concentration of 1.5 mM to the cultures. The samples were incubated in the plate reader for 2 hours.

2.1.2.3 Nutrient starvation

Nutrient deprivation was applied after growing NCTC 11168 to early stationary phase (OD_{600} $\approx$1.575 path-length corrected) and centrifugated for 3 minutes at 5000 x *g* before resuspension in sterile Ringer's solution (2.25 g/L NaCl, 0.08 g/L $CaCl_2 2H_2O$, 0.05 g/L $NaHCO_3$) in a 6-well plate (Greiner). The culture was incubated in the plate reader for 5 hours.

2.1.2.4 Acid shock

MH2 broth was calibrated to $\approx$pH 3.5 using 30 % Hydrochloric acid (Fisher Scientific) and filter sterilised with 0.22 μM pore membrane filters (Millipore). Once exponential phase was reached, the cultures were centrifugated for 3 minutes at 5000 x *g* before resuspending in the same volume of pH 3.5 MH2 broth and incubated for 10 minutes in the plate reader.

2.1.2.5 Anaerobic stress and heat shock

Once exponential phase was reached, NCTC 11168 culture in a 6-well plate (Greiner) was placed in a 55 °C incubator at normal atmospheric conditions for 3 minutes for heat shock. For anaerobic stress, NCTC 11168 culture in a 6-well plate (Greiner) was placed in a VAIN under anaerobic conditions (80 % N_2, 10 % CO_2, 10 % H_2) for 1 hour.

2.1.2.6 Peroxide stress

Once exponential phase was reached, 1.19 μL hydrogen peroxide (30% w/v) (Fisher Scientific) was added to 3.5 mL of NCTC 11168 culture in a 6-well plate (Greiner) for a final concentration of 3 mM. The culture was incubated for 10 minutes in the plate reader.

2.1.2.7 Collection of chicken juice

Whole frozen chicken (Waitrose, UK) was defrosted overnight at 4 °C and then at room temperature for a few hours. The chicken exudate was collected in 50 mL falcon tubes and filter sterilised using 0.22 μM pore membrane filters (Millipore). The chicken juice was stored at -20 °C up to one month until further use. At early stationary phase, 3.5 mL of NCTC 11168 culture was centrifugated at 5000 x *g* for 3 minutes and resuspended in the same volume of chicken juice in 25 cm^3 Vented Capped Tissue culture flasks (Falcon). 2.5 L AnaeroJar^{TM} (Oxoid) was pre-reduced to a microaerophilic environment using a CampyGen sachet (ThermoScientific). This was maintained each time the AnaeroJar was opened by using another sachet. NCTC 11168 resuspended in chicken juice was incubated in the microaerophilic AnaeroJar for 24 hours at 4 °C before cells were stabilized with cold killing buffer for RNA extraction. The same procedure was carried out with MH2 broth supplemented with 5 % chicken juice and MH2 broth only. The latter was used simultaneously as a control for the addition of chicken juice and as a cold stress.

2.2 Colony Forming Units (CFU)

Before and after the duration of each stress condition, 10 μL of culture was taken and diluted 10-fold using the Miles and Misra technique up to 10^{-8} dilution in a 96-well plate (Falcon) with phosphate buffered saline (PBS) (Miles *et al.*, 1938). 20 μL of 10^{-8} to 10^{-4} dilutions for three technical replicates for each of the three biological replicates were plated onto MH2 agar plates. The culture was dripped across the agar plate to spread the inoculum and incubated at 37 °C for two days before counting the CFUs.

2.3 RNA collection and extraction

0.5 volumes of cold killing buffer (20 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 20 mM NaN_3) stored at 4 °C was immediately added to stabilize 3.5 mL of bacterial culture from each well and spun down at 5000-10000 x *g* for 3-5 minutes. The pellets were snap frozen in liquid nitrogen and stored at -80 °C until further use. Sample pellets were placed on ice before extraction. Workspace and pipettes were cleaned with RNaseZap^{TM} (Sigma-Aldrich) prior to RNA extraction. Nuclease-free molecular biology grade water (Sigma-Aldrich), RNase-free filter tips, RNase-free eppendorf tubes (Sarstedt), and RNase-free 15 mL or 50 mL tubes (Falcon) were used throughout RNA extraction protocols. All centrifugation steps were carried out on Sigma 1-14K microcentrifuge (Sigma).

2.3.1 Hot-phenol

The hot-phenol protocol was adapted from the following studies: Blomberg *et al.* (1990), Jahn *et al.* (2008), and Dugar *et al.* (2013). All centrifugation steps for this protocol were carried out at 7200 x *g* at 4 °C. Cell pellets were resuspended in 600 μL of 0.5 mg/mL lysozyme in TE buffer (pH 8) pre-heated to 64 °C. 0.1 volume of 10 % SDS was added and vortexed gently. Samples were incubated at 64 °C for 2 minutes. 0.1 volume of 1 M sodium acetate was added. Then 1 volume of 65 °C hot-phenol was added to

the samples and incubated at 65 °C for 6 minutes with intermittent inversion every 40 seconds. Samples were incubated on ice for 5 minutes and centrifugated for 15 minutes. The upper aqueous phase was transferred to a new tube and 1 volume room temperature phenol was added and centrifugated as before. This was repeated once with phenol, then once with chloroform. 0.1 volume 3 M sodium acetate (pH 5.2), 1 mM EDTA, and 2 volumes of ethanol was added to the samples, mixed, and precipitated overnight at -80 °C.

Samples were centrifugated for 25 minutes. Pellets were washed with ice-cold 75 % ethanol and centrifugated for 5 minutes. The supernatant was removed and pellets left to dry for 20 minutes before being resuspended in 100 μL in nuclease-free water. Samples were stored in -80 °C until further use.

2.3.2 Bead-beating

400 μL of LETS buffer (0.1 M LiCl, 0.01 M Na_2EDTA, 0.01 M Tris-Cl (pH 7.4), 0.2 % SDS) was added to the sample pellet. 400 μL of phenol-chloroform-isomyl alcohol 25:24:1 was added to the sample and mixed up and down. The sample was then transferred to Lysing Matrix B, 0.1 mm tubes (MP biomedicals^{TM}). The tubes were placed in a FastPrep Homegenizer (MP biomedicals^{TM}) and bead-beat under the following settings:

Speed: 6.0 m/s
Adapter: CoolPrep
Time: 40 s
Lysing matrix: B
Quantity: 1 mg
Cycles: 3
Pause time: 300 s

The lysed cells were then centrifugated at top speed for 10 minutes in a Sigma 1-14k microcentrifuge (Sigma). The upper aqueous phase was then run through the RNA clean and concentrator^{TM} -25 kit (Zymo) according to the manufacturer's instructions. 15 μL of nuclease-free water was used to elute the RNA through the column twice for a total of 30 μL. The RNA was snap frozen in liquid nitrogen and stored at -80 °C until further use.

2.3.3 miRNAeasy Mini Kit

Modifications were made to the miRNAeasy Mini kit (Qiagen) manufacturer's manual for extraction in Gram-negative bacteria. Briefly, 10-20 μL of Proteinase K (Qiagen) and 1 mg/mL lysozyme in TE buffer was added to sample pellets to lyse the cells. The mixture was vortexed for 5 seconds and then incubated at room temperature for 5 minutes with intermittent vortexing. 1 mL of Qiazol (Qiagen) was added and mixed by vortexing for 3 minutes then incubated for 5 minutes. 200 μL of chloroform was added, shaken, and incubated at room temperature for 3 minutes. The upper aqueous phase was transferred to a new tube and added to 1.5 volumes of 100 % ethanol and mixed. The samples were then transferred to a silica spin

column (provided with the kit) and centrifugated at 8000 x *g* for 15 seconds at room temperature.

DNase treatment was carried out twice through the spin column using DNase I (Qiagen). 350 μL of RWT buffer (provided with the kit) was used to wash the spin column. 70 μL of buffer RDD (provided with the kit) was added to 10 μL of DNase I and the entire volume (80 μL) was pipetted directly onto the spin column and incubated for 15 minutes at room temperature. The spin column was once again washed with 150 μL of RWT buffer and the whole process was repeated once. Then 500 μL RPE buffer (provided with the kit) was used to wash the spin column. The spin column was transferred to a new tube and 17.5 μL of nuclease-free water was added to the tube and spun down. This was repeated for a total of two times for a final volume of 30 μL.

2.3.4 Modified Bead-beating protocol

This method is an improvement on the previous bead-beating protocol (Section 2.3.2) and the final optimised RNA extraction method used for sequencing. RNA samples from all the conditions were batch extracted to prevent batch effects.

600 μL of LETS buffer (as described before) was added to sample pellets and transferred to Lysing Matrix B, 0.1 mm tubes (MP biomedicals^{TM}). The tubes were placed in a FastPrep Homegenizer (MP biomedicals^{TM}) and bead-beated as previously described. The samples were then centrifugated at 13000 x *g* for 10 minutes at 4 °C. 1 volume of 125:24:1 phenol:chloroform:isomyl alcohol was added to the mixture. The samples were then shaken at 1400 rpm for 5 minutes at room temperature before centrifugation at top speed. The upper phase was transferred to a new tube and 1 volume of 125:24:1 Phenol:chloroform:isomyl alcohol was added once again and shaken and centrifugated as before. The upper phase was transferred to a new tube and 0.1 volume of 3 M sodium acetate pH 5.2 and 1 volume isopropanol was added then mixed by inverting several times. The samples were precipitated at -20 °C overnight.

Samples were centrifugated at top speed at 4 °C for 15 minutes. The supernatant was removed and pellets washed (without resuspending) with freshly made 70 % ethanol. The samples were centrifugated once again as described before and supernatant removed. Pellets were left to dry for 15 minutes and then resuspended in 50 μL of RNase-free water) and incubated on ice for three hours, then 30 minutes at room temperature. RNA samples were then snap frozen in liquid nitrogen and stored at -80 °C until further use.

2.4 DNA and rRNA removal

The concentration of RNA and DNA was measured using Qubit® 2.0 Fluorometer assay kit (Life Technologies, UK) according to the manufacturer's guide.

2.4.1 DNase treatment

On column DNase treatment was incorporated in the miRNeasy Mini kit (Qiagen) protocol as per manufacturer's instructions (Section 2.3.3), but for the hot-phenol and bead-beating methods the following in-solution protocol was used for DNA removal. Genomic DNA contamination was removed from 2 x 15 μg RNA samples using TurboDNase (Ambion) according to the manufacturer's protocol, although DNase inactivation reagent was not used to avoid downstream effects. 1 μL SUPERase In™ RNase Inhibitor (Invitrogen) was added per 50 μL sample to prevent RNA degradation. Samples that underwent miRNAeasy Mini Kit extraction used DNaseI (Qiagen) instead as part of its protocol (see Section 2.3.3). After DNase treatment samples were purified using RNA Clean and concentrator™ -5 kit (Zymo) according to the manufacturer's instructions. All samples were checked for residual genomic DNA with bacterial forward primer Bac27F and universal reverse primer Univ1492R against the 16S rRNA gene using GoTaq Green master mix (Promega).

Bac27F - (5'-AGAGTTTGGATCMTGGCTCAG-3')
Univ1492R - (5'-CGGTTACCTTGTTACGACTT-3').

If the RNA samples still had DNA contamination after 30 cycles of PCR amplification a further round of DNase treatment was used, although this reduces the RNA concentration. Once samples were free from the DNA PCR product, RNA integrity and further confirmation of DNA removal was checked by running the Agilent 2100 Bioanalyzer RNA 6000 Pico kit (Agilent, UK) according to the manufacturer's guide.

2.4.2 Ribosomal RNA depletion

Ribosomal RNA (rRNA) was removed from the samples using Ribo-Zero™ rRNA removal kit for bacteria (Illumina, UK) according to manufacturer's protocol. Samples were once again run on Agilent 2100 Bioanalyzer RNA 6000 Pico kit (Agilent, UK) according to the manufacturer's guide to confirm complete removal of rRNA.

2.5 Library preparation

Depending on the purpose of the sequencing protocol different library preparation methods were used. After cDNA libraries were synthesised DNA concentration was measured using Qubit^{TM} High sensitivity DNA kit (Life technologies, UK). cDNA libraries were checked for the fragment size on Agilent 2100 Bioanalyzer DNA High-sensitivity kit (Agilent, UK) before proceeding with RNA-sequencing.

2.5.1 Standard Illumina TruSeq library preparation

The Illumina TruSeq stranded mRNA RNA library preparation kit was used to generate cDNA libraries from rRNA depleted RNA according to the manufacturer's guide but omitting the initial polyA capture step. For comparing different RNA extraction methods 65 μL of AMPure beads XP (Agencourt) were used for the final resuspension step rather than the standard 50 μL. All in-line control reagents were used.

2.5.2 Multiplex (RNAtag-Seq) library preparation

This protocol was adapted from Shishkin *et al.* (2015) and visually represented in Figure 2.1. Briefly, ten adapters with a unique 8 bp barcode sequence were ligated to each RNA sample at the 3'-OH end. Then tagged RNA samples were pooled in batches of ten so that each pool did not have duplicate tags. Each pool underwent Ribo-Zero^{TM} rRNA depletion before first strand cDNA synthesis. An adapter complementary to the P7 sequence was ligated to the 3'-OH end forming the basis for PCR enrichment using P7 primers containing the i7 index and a universal P5 primer. A different P7 primer was used for each pool so that each sample has a unique pair of barcode sequence and i7 index. Libraries were sequenced on the NextSeq 550 system on a 150 cycle High-Output cartridge (2 x 75) over two runs. The first run had 30 samples (10 conditions, three biological replicates each) in combination with *B. subtilis* samples prepared the same way. The second run was standalone NCTC 11168 samples with 42 samples (all replicates of the remaining 12 conditions and two additional conditions that were repeated).

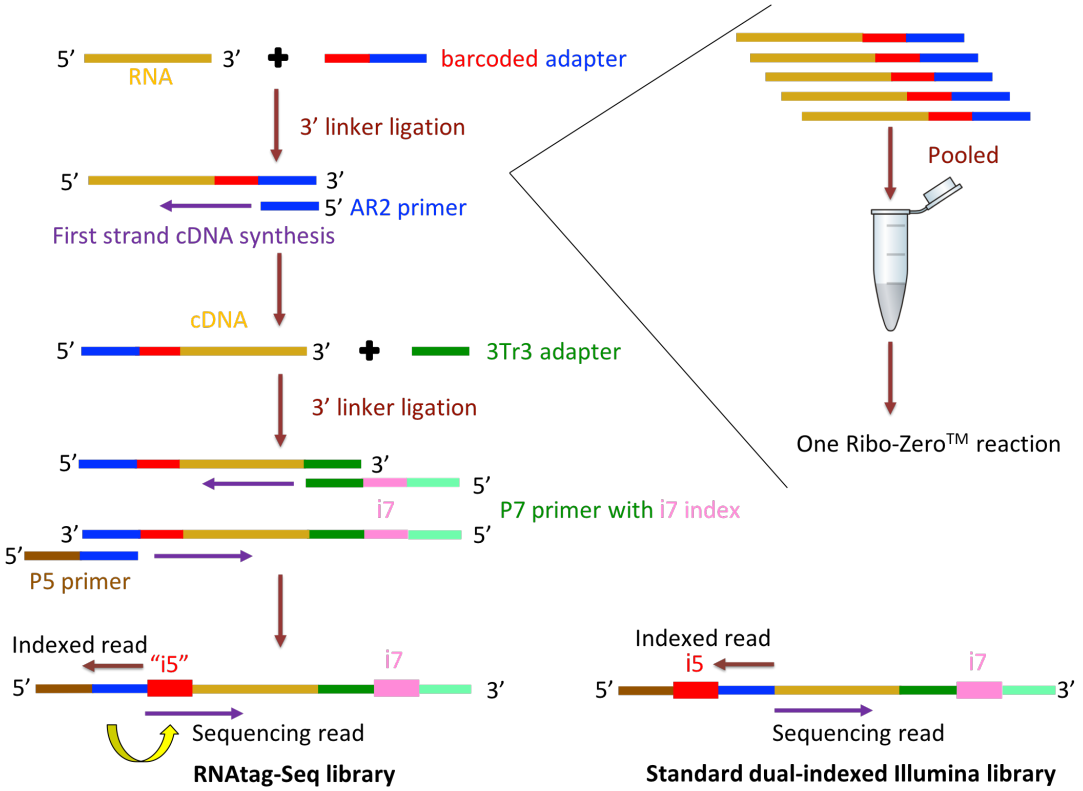


Figure 2.1: Schematic diagram of RNAtag-Seq protocol. RNA samples were tagged with a unique barcoded adapter for each sample in batches of 10 before pooling together and undergoing a single ribosomal depletion reaction. The pooled samples then underwent cDNA synthesis and adapter ligation before PCR enrichment. Adapted from Shishkin *et al.* (2015)

2.5.3 Cappable-seq library preparation

This protocol was adapted from Ettwiller *et al.* (2016) and from New England biolabs (NEB) website [1] as depicted in Figure 2.2. All RNA samples (three biological replicates) were diluted to 100 ng/μL$\pm$12% and pooled together prior to Cappable-seq processing. Two of these biological replicates from each condition were pooled together for the Cappable-seq protocol. A non-enriched control library was prepared by omitting the Streptavidin capture step and instead cleaned and purified with RNA Clean and concentrator^{TM} -5 kit (Zymo).

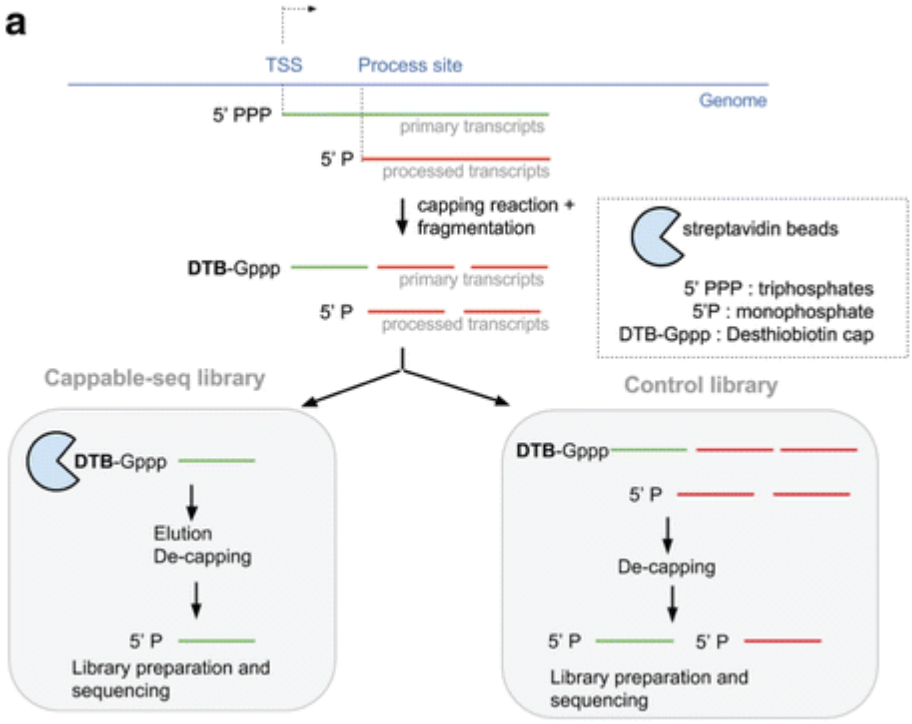


Figure 2.2: Schematic diagram of Cappable-seq protocol. 5'-PPP ends (primary transcripts) of RNA samples were capped with 3'-Desthiobiotin-GTP (DTB-GTP). The treated sample was enriched using streptavidin beads whereas the control sample omits this step. The samples were then de-capped and prepared for sequencing. Taken from Ettwiller *et al.* (2016).

[1] `https://international.neb.com/protocols/2018/01/19/cappable-seq-for-prokaryotic-transcription-start-site-determination`

Briefly, 3 μg of pooled RNA (for both control and treated) were capped with 3'-Desthiobiotin-GTP (DTB-GTP) (NEB) using the Vaccinia Capping system (NEB) according to the manufacturer's protocol. S-Adenosylmethionine (SAM) was not included in the reaction as capping is more efficient in the absence of SAM (NEB). RNA samples were cleaned and purified using RNA Clean and concentrator^{TM} -5 kit (Zymo). Instead of the >200 nt cut-off stated in the protocol, the normal total RNA protocol from Zymo was used with a total of 4 washes. A stock of 20 mM D-biotin was prepared in 1 M Tris pH 7.5 and heated to 65 °C before storage at 4 °C. Capped RNA was fragmented using T4 polynucleotide kinase (NEB) and enriched using hydrophilic streptavidin magnetic beads (NEB). The sample was incubated with D-biotin diluted to 1mM in 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA then cleaned and eluted with AMPure XP beads (Agencourt) as stated in the protocol. The streptavidin capture step and subsequent RNA clean up was repeated once more before decapping of DTB-GTP using 10X Thermopol Buffer (NEB) and RNA 5´ Pyrophosphohydrolase (RppH) (NEB) according to the manufacturer's guide. Samples were stored at -80 °C until further use.

The processed RNA was then generated into sequencing libraries using NEBNext Small RNA Library Prep Set for Illumina (NEB). 12 cycles of PCR amplification was used for the control sample and 18 cycles of PCR amplification were used for the enriched sample. Samples were cleaned with the Monarch® PCR & DNA Cleanup kit (5 μg) (NEB) and bioanalyzed to check for overamplification and adapter primer dimers. 0.9X AMPure XP beads (Agencourt) were used for size selection and the cDNA libraries were checked once more on the Bioanalyzer before sequencing on the Illumina MiSeq using a 150 cycle v3 cartridge (Illumina) with 5 % phiX control.

2.6 RNA-sequencing

2.6.1 Denaturing and diluting libraries

cDNA libraries were diluted to either 2 nM or 4 nM pool calculated from cDNA concentration and average fragment size. Libraries were denatured according to the appropriate system in "Denature and dilute libraries guide" for: MiSeq[2] or NextSeq[3] (Illumina, UK). 9-9.5 pM and 1.6 pM was used as the final concentration respectively to prevent overclustering. For the MiSeq test run using RNAtag-Seq libraries, 5 μL 200 mM Tris-HCl (pH 7) was also added to neutralise the NaOH step as is recommended for the NextSeq protocol. 5 % PhiX control was added for TruSeq libraries and 20 % PhiX for RNAtag-Seq libraries.

2.6.2 Loading Illumina MiSeq and NextSeq

Both instruments Illumina MiSeq and NextSeq were loaded according to the manufacturer's guide. 2 x 76 cycles (one extra cycle for phasing) for paired-end sequencing was used for RNAtag-Seq and normal RNA-seq. Cappable-seq used 1 x 100 cycles single-end sequencing but with no adapter trimming as NEB indexes were used in lieu of Illumina indexes.

2.7 Data analysis

All data was processed using command-line on a virtual machine on CLIMB (the Cloud Infrastructure for Microbial Bioinformatics) (Guest *et al.*, 2016). The MiSeq instrument automatically converts binary base call (BCL) files to fastq files. However, the NextSeq does not convert BCL files so this was done manually. The MiSeq was unable to convert BCL files automatically using RNAtag-Seq libraries due to the nature of the barcoding system used, thus the conversion was also done manually.

2.7.1 De-multiplexing RNAtag-Seq data

BCL files from RNAtag-Seq runs were converted to fastq files and de-multiplexed according to the i7 index using the software bcl2fastq v2.20.0.422 (Illumina). The input for the software required a sample sheet comma separated value (csv) format with the i7 sequences placed in the run directory. RunInfo.xml (found in the run directory) was edited so that Indexed Read 3 was changed to 'N' indicating the i5 index will not be read. When de-multiplexing data obtained from the NextSeq run, the options `--no-lane-splitting` was included as there are four lanes on the NextSeq flowcell, so fastq files were not split according to lane. The following command was used:

fastq {--no-lane-splitting} -R [run directory] -o [output directory]

A Python script named Example_tile.py written by Leonidas Souliotis was then used to de-multiplex the fastq files according to the unique tag/barcode which has been incorporated as part of the first read Read 1.

[2] `http://emea.support.illumina.com/downloads/prepare_libraries_for_sequencing_miseq_15039740.html`
[3] `http://emea.support.illumina.com/downloads/nextseq-500-denaturing-diluting-libraries-15048776.html`

The tag sequences were inputted into the dictionary list in the script before running the script. The script automatically searches for all fastq files in the directory and therefore does not require an input.

2.7.2 Analysis of RNAtag-Seq data

General feature format (GFF) and fasta files of NCTC 11168 were downloaded from NCBI accession number AL111168 AL139074-AL139079 and used for data analysis. As repeat regions were not annotated in the 'feature' column of the GFF file (3rd column) this was rectified. Published ncRNAs from Dugar *et al.* (2013) and Porcelli *et al.* (2013) and putative novel ncRNAs identified from Section 2.7.4 were added to the GFF file prior to analysis. Duplicate values within a few bp from the studies were combined taking the earliest start and latest end position before adding to the GFF file, but transcripts with the same start site but with differing lengths were included from both studies with the source noted. A reference genome was created from the fasta file using bowtie2-build (Langmead and Salzberg, 2012). Fastq files from RNAtag-Seq runs had the first 9 bp trimmed from Read 1 to remove the unique tag/barcode before processing of reads. Read 1 was reverse complemented using the Seqtk[4] tool so that Read 1 was in the correct orientation (5'-3').

Reads were mapped to the reference genome using bowtie2 generating sequence alignment map (SAM) files with the parameters `--very-sensitive-local` and `--ff` indicating a preference for accuracy over time and the orientation of the strands respectively (Langmead and Salzberg, 2012). SAM files were converted to binary alignment map (BAM) files, and indexed and sorted using SAMtools (Li *et al.*, 2009). Mapped and unmapped reads were filtered from the BAM files and only mapped reads were subsequently used for further analysis. CoverageBed from Bedtools v2.27.1 was used to calculate the coverage of reads mapping to each gene feature (Quinlan and Hall, 2010). The output from coverageBed was then filtered for the read counts according to "genes", "ncRNAs", and "pseudogenes" using an edited script PrepareDESeq_gff3_locus.py originally written by Richard Brown (available in supplementary materials[5]). The filtered text output files were used as input for DESeq2 for normalisation and to call differentially expressed genes (Love *et al.*, 2014). DESeq2 was run on R Studio (R Core Team, 2018).

Another normalisation method used to account for gene length and depth was the transcripts per million (TPM) approach by Wagner *et al.* (2012). For each gene, the RPK (reads per kilobase) was calculated by dividing the read counts for each sample by the gene length. The scaling factor was derived by taking the sum of RPK values per sample for each gene and dividing by 1 million. The RPK is then divided by the scaling factor. Functional enrichment analysis was done using the STRINGdb package on R (v10) (Franceschini *et al.*, 2013; Szklarczyk *et al.*, 2015). A combined score of 900 was used as a cut-off for interactions so that only high confident interactions were called. Heat maps and plots were made using the pheatmap package on R.

[4] `https://github.com/lh3/seqtk`
[5] `https://github.com/Jenna-Lam/Supplementary-material-for-thesis`

2.7.3 Analysis of Cappable-seq data

Scripts used for analysing data from the Cappable-seq run were git cloned (downloaded) from the Git Hub repository[6] used in Ettwiller *et al.* (2016)'s study. The 3' adapter was trimmed manually using Cutadapt version 1.18 with default parameters and the input `-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC` as was used in Ettwiller *et al.* (2016) (Martin, 1994). Reads were mapped and aligned to the reference genome using bowtie2 with the parameters `--very-sensitive` and `-L 16`. Conversion to sorted bam files were generated as previously in Section 2.7.2. Perl scripts from the Git Hub repository were ran with the default parameters to give output files which were used for analysis in Chapter 4:

perl bam2firstbasegtf.pl --bam control_library.bam --cutoff 0 --out control.gtf
perl bam2firstbasegtf.pl --bam Cappable-seq_library.bam --cutoff 1.5 --out enriched.gtf
perl filter_tss.pl --tss enriched/gtf --control control.gtf --cutoff 0 --out TSS_enriched.gtf
perl cluster_tss.pl --tss TSS_enriched.gtf --cutoff 10 --out TSS_enriched_cluster_5.gtf

That is cut-off 0 for the control sample, cut-off 1.5 relative read score (RRS) for the enriched sample and cut-off 0 for the filtering. 10 bp was used for clustering TSS together.

Transformation of Nucleotide Enrichment Ratios (ToNER) was used as a statistical tool to determine significantly enriched TSS positions (Promworn *et al.*, 2017)[7]. The options `-r start` specifying to only use the first position at the 5' end, `-q 0.8` to specify the R-square value, and `-g start` to look at annotations upstream of genes from a GFF input file were used.

To search for nucleotide bias and therefore promoter motifs upstream of TSS, the tool Multiple Em for Motif Elicitation (MEME) version 5.0.5 was used[8]. A fasta file containing sequences 50 nt upstream of each identified TSS from Cappable-seq was submitted to MEME with the option `search given strand only` selected.

All custom made scripts used to analyse identified TSS from Cappable-seq for categorisation of TSS type, sigma factor preference, and generation of the Artemis file containing promoters are available in the supplementary materials[9] (Section Appendix). To determine conditionally expressed TSS, 10 nt of each TSS transcript was assigned a name and collated into a GFF file format. This GFF file was used as input for coverageBed against bam files generated from RNAtag-Seq for each individual replicate per condition, and TPM values were calculated as above for each read count. TPM values ≤ 10 were treated as not expressed. The method was adapted from Kröger *et al.* (2013).

[6] https://github.com/Ettwiller/TSS
[7] http://www4a.biotec.or.th/GI/tools/toner
[8] http://meme-suite.org/tools/meme
[9] https://github.com/Jenna-Lam/Supplementary-material-for-thesis

2.7.4 Identification of ncRNAs

ncRNAs were identified using The_toRNAdo_script.py[10] (Hermansen *et al.*, 2018). Before running the script it is necessary to generate input files for the toRNAdo script using Prepare_for_toRNAdo_v2.py[11] in the same directory as the sorted BAM files. Version 2 (v2) of the script was used for versions of Bedtools 2.24.0 or above. This generates output directories containing input files for toRNAdo. After running toRNAdo the output can be found in the directory `tornado_output` containing a csv file listing the start and stop positions, strand information, and type of putative ncRNA found. Wig files are also available for optional visualisation on the Artemis Genome browser. A modified GFF file with no sequence features was used as an input. Putative ncRNAs were compared with TSS determined in Section 2.7.3 and novel ncRNAs with a TSS within 10 bp were added to the GFF file before using coverageBed to determine read counts. ncRNAs that were within 50 nt at both the start and end positions were combined together and noted in the GFF file. Thus, ncRNAs were included when normalising for TPM and for DESeq2 analysis.

[10] `https://github.com/pavsaz/toRNAdo`
[11] `https://github.com/pavsaz/Prepare_for_toRNAdo`

Chapter 3

Optimisation of RNA extraction methods and stress conditions

3.1 Introduction

C. jejuni is known to be a fastidious microorganism to culture in the laboratory due to its relatively slow growing nature and microaerophilic growth requirements. This is further exacerbated by the VBNC state where cells change morphology and become coccoid shaped in late stationary phase (Rollins and Colwell, 1986). This can lead to clumping which affects OD readings. Optimisation of growth curves is required for characterising the growth of *C. jejuni* and to select points for harvesting cells for RNA extraction. Any additional supplements and/or change of media will affect growth and therefore needs to be characterised.

Transcriptional activity is occurring constantly at any given time in vegetative growing cells. This is regulated by a number of transcription factors to initiate transcription and therefore gene expression. Any changes in the environment may require a rapid response to alter gene expression accordingly. In bacteria the SOS or stringent response can be activated during severe stresses or aged cultures, however, *C. jejuni* does not possess the canonical stringent responses other Gram-negative bacteria typically exhibit during late stationary phase (Kelly *et al.*, 2001). In this project, we are interested in investigating how NCTC 11168 adapts to stress. If the applied stress is too harsh, this can lead to cell lysis and death resulting in an increase in RNA degradation affecting RNA integrity. RNase R activity increases drastically under many stress conditions as the enzyme becomes more stable, as was shown in *E. coli* (Deutscher, 2015). Hence, it is crucial to ensure that any stresses applied for experimental procedures are not causing major cell death so that results from RNA-seq are not misrepresented.

The integrity of RNA obtained from RNA extraction can highly affect the output of transcriptomic data, reducing the accuracy of transcript quantification, which in turn can lead to ambiguous differential gene expression results from RNA-seq (Reiman *et al.*, 2018). Thus, an RNA extraction method that yields high quality RNA and is able to retrieve the highest amount of RNA from bacterial cells is desired. An arbitrary way to measure integrity is by using the RNA integrity number (RIN) value from Bioanalyzer traces, showing the quality of 16S and 23S rRNA peaks. The traces are also useful for checking DNA contamination and whether rRNA depletion was successful.

Putative ncRNAs (which have the potential to be sRNAs) are commonly discovered in transcriptomic data and are increasing in relevance as post-transcriptional regulators. Therefore, an RNA extraction method which does not compromise small transcripts will be beneficial for investigating sRNAs and the transcriptional landscape in more detail. Many commercial kits are available on the market for RNA extraction and are normally easier to use and less labour intensive, although a major downside is they can be expensive. One concern is that kits normally use columns to purify RNA and carry the risk of losing sRNAs which are typically shorter than mRNA transcripts. A more traditional way to extract RNA is using the hot-phenol method, generally known for extracting good quality RNA. However, hot-phenol is more dangerous than commercial kits and is not as convenient to carry out. RNA extraction is a crucial methodology used exhaustively in this project, so a careful decision of which protocol to use would be of great benefit. For the purposes of this project three RNA extraction methods were compared qualitatively and quantitatively.

Chapter Aims:

- Optimise the growth of NCTC 11168 under standard laboratory conditions and conditions with supplements
- Ensure the severity of stress shocks are bacteriostatic to avoid major cell death
- Compare RNA extraction methods and choose the best method to retain ncRNAs

3.2 Optimisation of growth curves and stress conditions

Growth of NCTC 11168 at 37 °C was optimised in the FLUOStar Omega microplate reader (BMG Labtech) with an ACU in 6-well plates (Greiner). *C. jejuni* can be grown in the VAIN in 175 cm^2 Vented capped tissue culture Flasks (Falcon) which is the conventional way of culturing *C. jejuni*, however, there were many impracticalities using this method. Previous attempts to grow 50 mL of NCTC 11168 culture in the VAIN proved difficult; measuring OD_{600} every two hours over a 48-hour period was not practical. There were also issues with the growth stalling, although this was overcome by inoculating more cells in the overnight cultures. Growth of *C. jejuni* has been shown to be feasible in a 96-well plate in a microplate reader with an ACU (Haigh and Ketley, 2011). Using a microplate reader provides very reproducible results that can measure intervals of 30 minutes producing more accurate and fine tuned growth curves with easier temperature adjustments for application of stress conditions.

3.2.1 Growth of NCTC 11168 in standard conditions in Mueller-Hinton media

Standard laboratory growth conditions for *C. jejuni* are either carried out at 37 °C or 42 °C in microaerophilic conditions. 37 °C is the body temperature of humans and other mammals, whereas 42 °C is the body temperature of chicken and other avian species. Both host temperatures are within the range of optimal growth for *C. jejuni*. 37 °C is the standard growth temperature used in our laboratory. NCTC 11168 was inoculated and grown in 3.5 mL MH broth in a 6-well plate (Greiner) in the microplate reader using 6 % CO_2 and 4 % O_2 and controlled by an ACU shaking at 200 rpm. The two left hand-side wells were used as blanks and the rest of the wells were used for each overnight replicate. The cultures in each well are considered as pseudo-independent as they were derived from separate agar plates but are referred to as technical replicates since they were carried out in the same environment at the same time. 3.5 mL was found to be the optimal volume in the well without any cross contamination (data not shown). Figure 3.1 shows the average $\log_{10}$ of three independent standard growth curves from different days i.e. biological replicates of NCTC 11168 using MH broth (Oxoid), so in total an average of 12 data points. The small error bars indicate that the growth curves are highly reproducible with little variance.

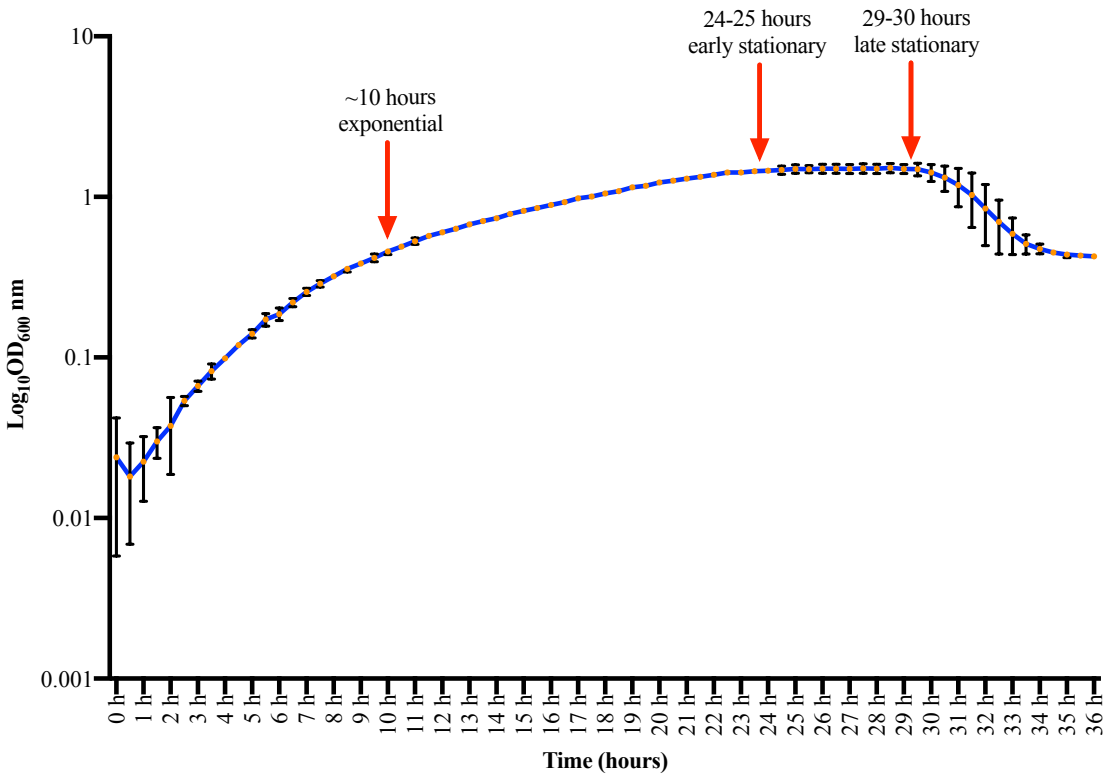


Figure 3.1: Standard $\log_{10}$ growth curve of NCTC 11168 in MH broth. An overall average of 12 overnight replicates derived from three independent experiments that were carried out on different days, with four technical replicates each. N.B. One entry was omitted as it was an anomaly. The standard deviation error bars were calculated from the three independent experiments. Red arrows indicate selected time points chosen for RNA extraction.

There was a discrepancy between measuring OD_{600} using the spectrophotometer and the microplate reader as the instrument was not able to automatically apply path-length correction to a volume of 3.5 mL in the wells. Different concentrations of overnight cultures were measured and compared between the two instruments to find the fold-change and determine the path-length correction (Figure 3.2). The correction needed to be applied was a 4.5-fold increase in OD_{600} when using the microplate reader compared with the spectrophotometer, calculated by 1/slope. For example, an OD_{600} 0.3 measured on the microplate reader is the equivalent of $\approx$1.35 on the spectrophotometer. When 2.5 mL was used for comparing RNA extraction methods (see Section 3.3), the appropriate correction was a 10-fold increase (Figure 3.3). Henceforth, all reported OD readings and data points visually represented in Figures 3.1 and 3.4-3.11 are path-length corrected.

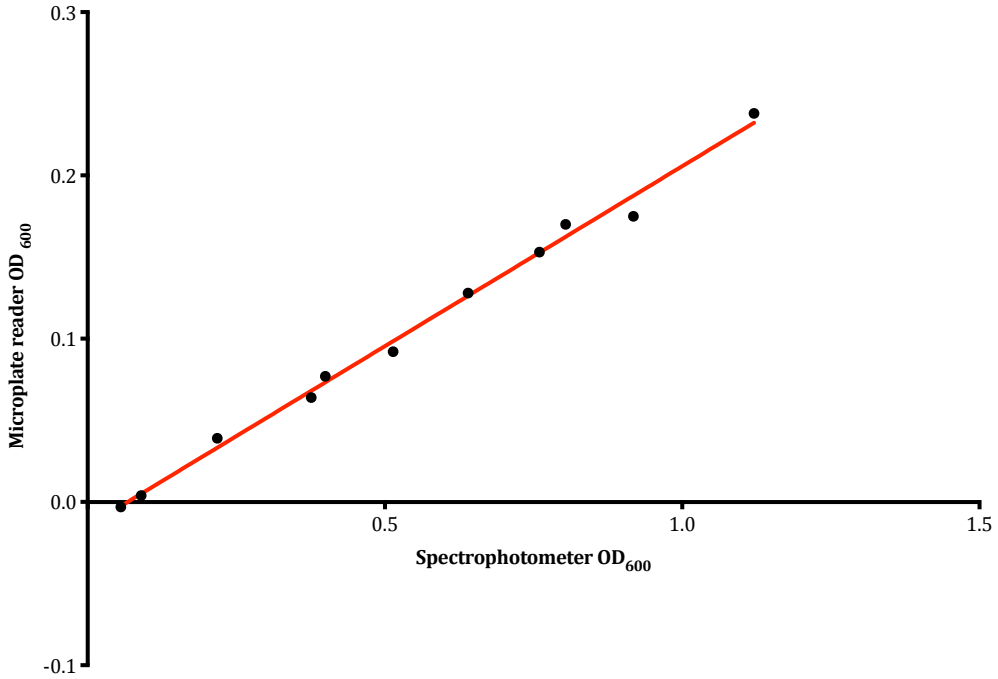


Figure 3.2: A linear regression graph of the ratio between the spectrophotometer and the microplate reader for 3.5 mL volume. The line of best fit is in red. There is approximately a 4.5-fold change from the microplate reader to the spectrophotometer.

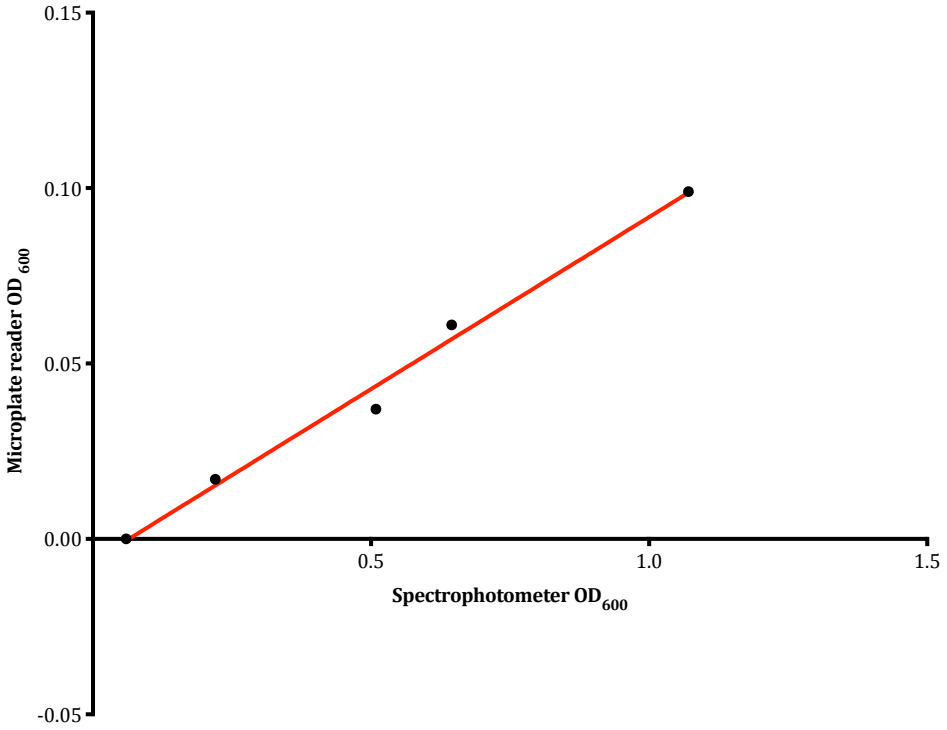


Figure 3.3: A linear regression graph of the ratio between the spectrophotometer and the microplate reader for 2.5 mL volume. The line of best fit is in red. There is approximately 10-fold change from the microplate reader to the spectrophotometer.

3.2.2 Problems with media used

Batches of MH broth (Oxoid) used for the initial experiments were recalled by the supplier as there were problems with the levels of magnesium, leading to precipitation in the media and unavailability from the supplier. As this stalled experimental work and due to time constraints subsequent experiments used MH2 cation-adjusted broth (Sigma), which is a more standardised version of MH broth as each batch of media should have equal levels of cations, reducing batch variability. Different brands of MH broth have differing levels of cation concentrations and this was shown to affect antimicrobial susceptibility when testing for minimum inhibitory concentrations (MICs) (Girardello *et al.*, 2012b). To no surprise there was a difference in growth between MH (Oxoid) and MH2 cation-adjusted (Sigma) broth (Figure 3.4). Growth curves of NCTC 11168 were characterised once more in MH2 broth but repeated with only two independent experiments due to compromise of lost time, unavailability of the plate reader, and previous data having shown to be very reproducible. Different stages of growth phases were chosen for RNA extraction as seen in Figure 3.5: exponential phase OD_{600} 0.765-0.9, early stationary phase OD_{600} ≈1.575 and above, and late stationary phase before decline ≈38 hours. Red arrows indicate the approximate time taken to reach the desired range of OD_{600}.

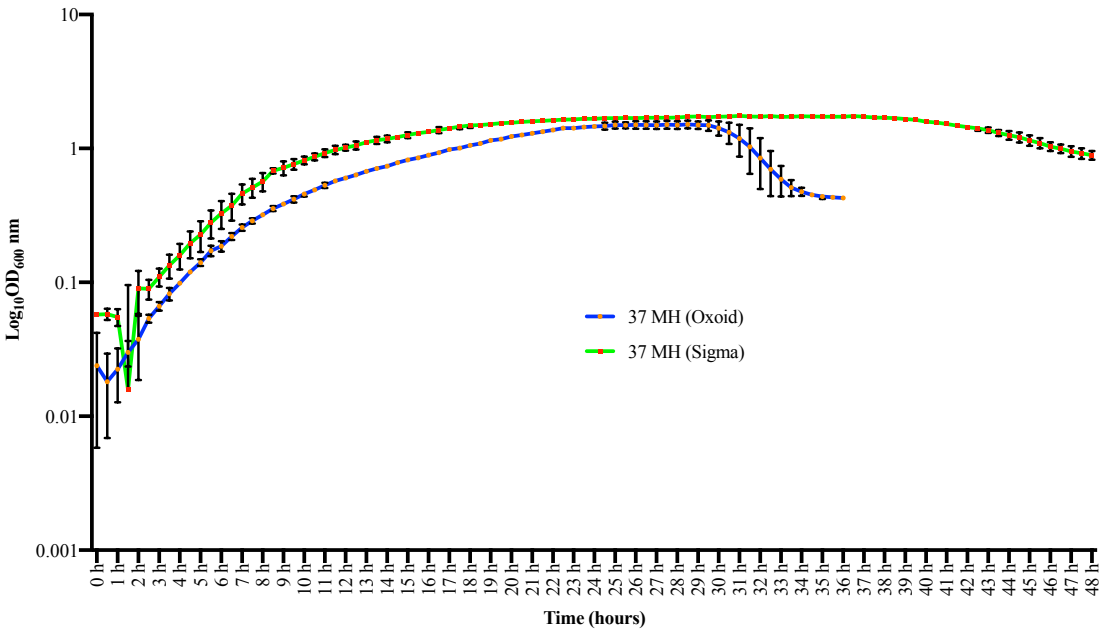


Figure 3.4: Standard $\log_{10}$ growth curves of NCTC 11168 using MH broth (Oxoid) and MH2 cation-adjusted broth (Sigma-Aldrich). Three independent experiments using MH broth and two independent experiments using MH2 both were carried out on different days with an average of four overnight technical replicates each originating from separate agar plates. The standard deviation error bars were calculated for the independent experiments.

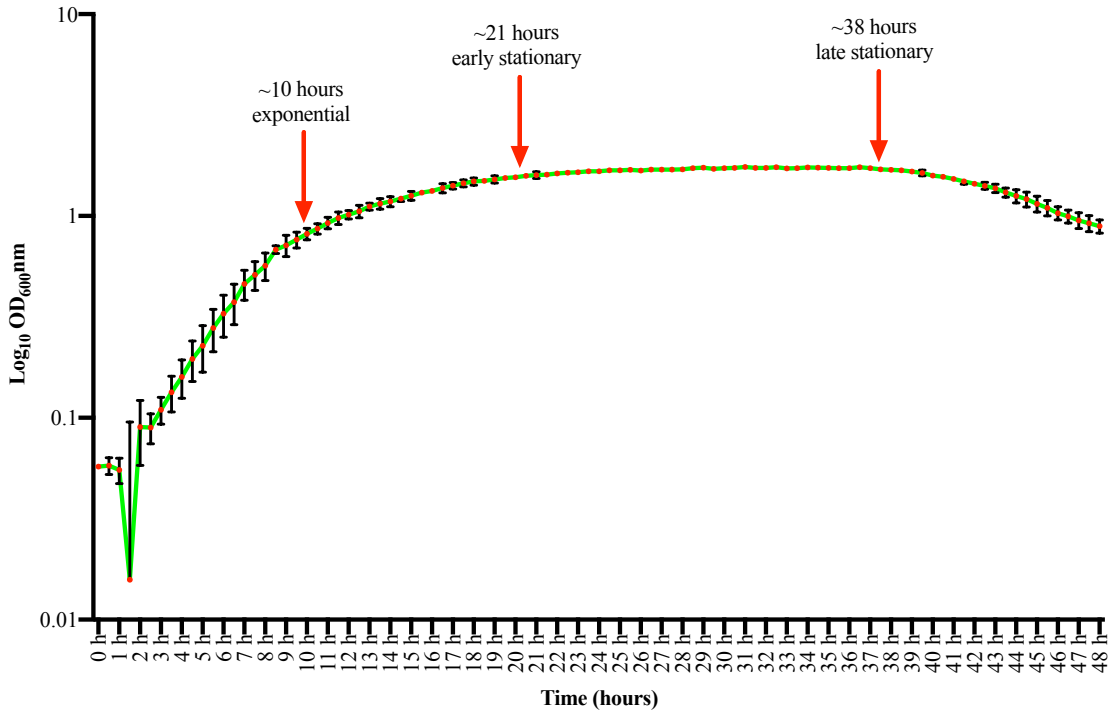


Figure 3.5: Standard $\log_{10}$ growth curve of NCTC 11168 in MH2 broth (Sigma-Aldrich). With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments. Red arrows indicate selected time points chosen for RNA extraction.

3.2.3 Growth curves at 42 °C

42 °C is the chicken body temperature and many laboratories use this for standard growth and maintenance of *C. jejuni* (Figure 3.6). There are, however, differences in *in vitro* growth between 37 °C and 42 °C as shown in Figure 3.7. RNA was collected at exponential OD_{600} 0.675-0.81, early stationary OD_{600} 1.58-1.71, and late stationary around 26 hours when growth has plateaued but before the decline phase (Figure 3.6). Red arrows indicate the approximate time taken to reach the desired range of OD_{600}.

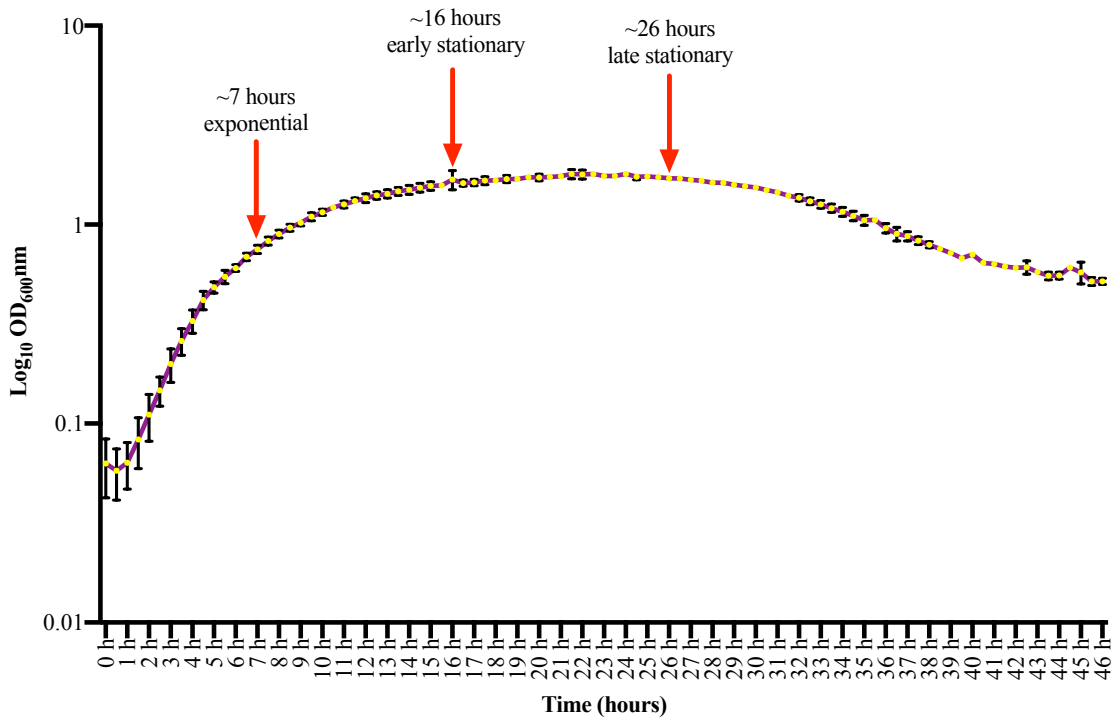


Figure 3.6: Log_{10} growth curve of NCTC 11168 in MH2 broth at 42 °C. With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments. Red arrows indicate selected time points chosen for RNA extraction.

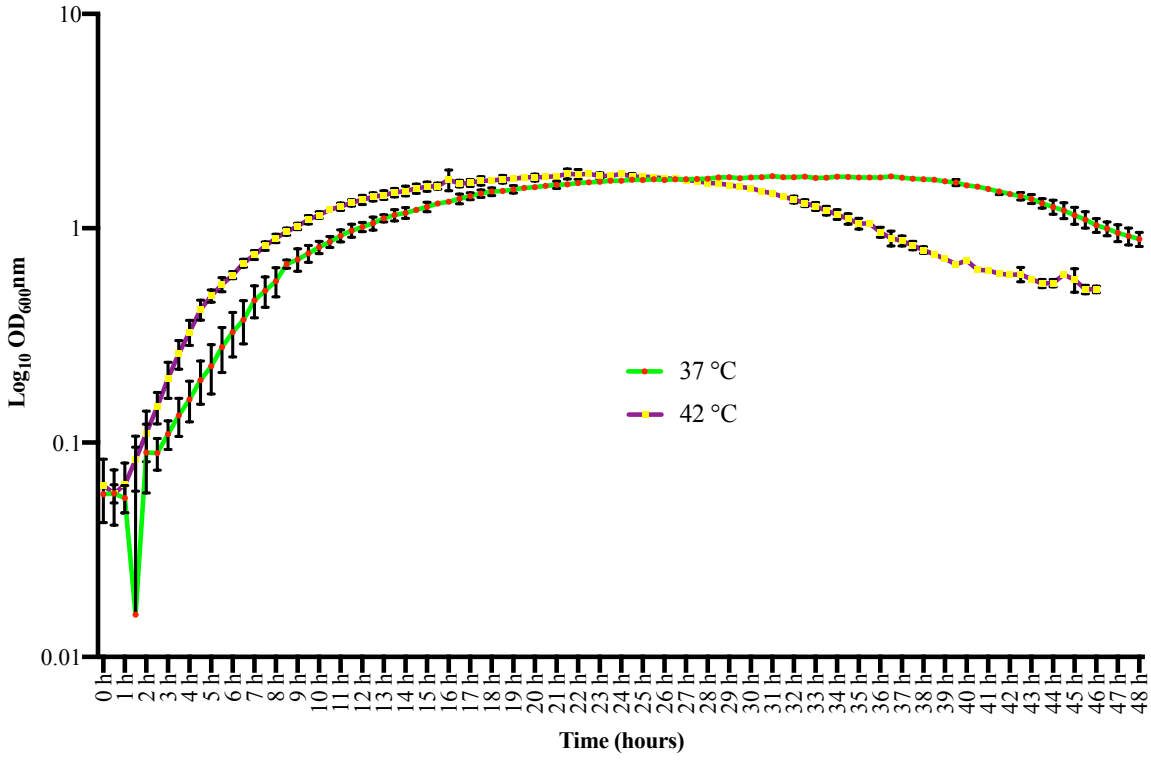


Figure 3.7: Log_{10} growth curve of NCTC 11168 at 37 °C vs 42 °C. With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments.

3.2.4 Sodium deoxycholate stress

Growing NCTC 11168 in 0.1 % sodium deoxycholate, a bile salt, mimics the stress *C. jejuni* encounters when ingested in the human body. The stress was adapted from Malik-Kale *et al.* (2008) and hindered the growth of NCTC 11168 as compared to standard routine growth conditions as demonstrated in Figure 3.8. Figure 3.9 shows the $\log_{10}$ growth curve and the points of RNA extraction: OD_{600} 0.076-0.135 for exponential phase and OD_{600} 0.27-0.34 for early stationary phase. The pellet collected at early stationary phase also showed a loss of pigmentation, losing the typical pink colour of *Campylobacter* cells.

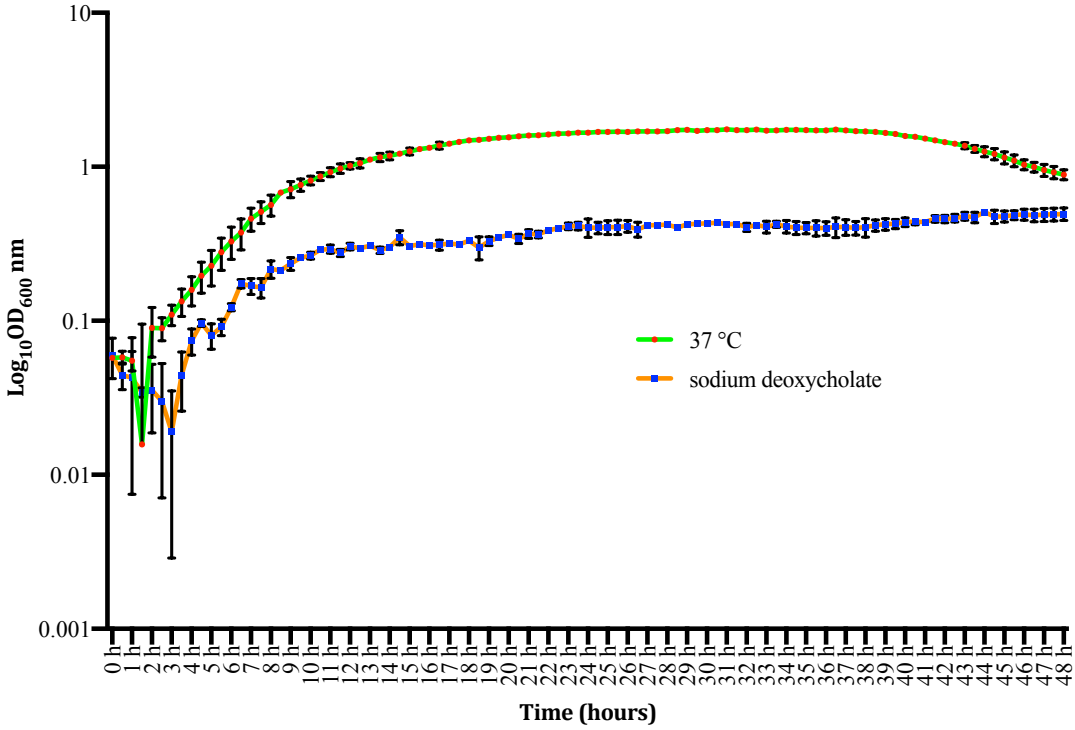


Figure 3.8: Log_{10} growth curve of NCTC 11168 at 37 °C vs addition of 0.1 % sodium deoxycholate. With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments.

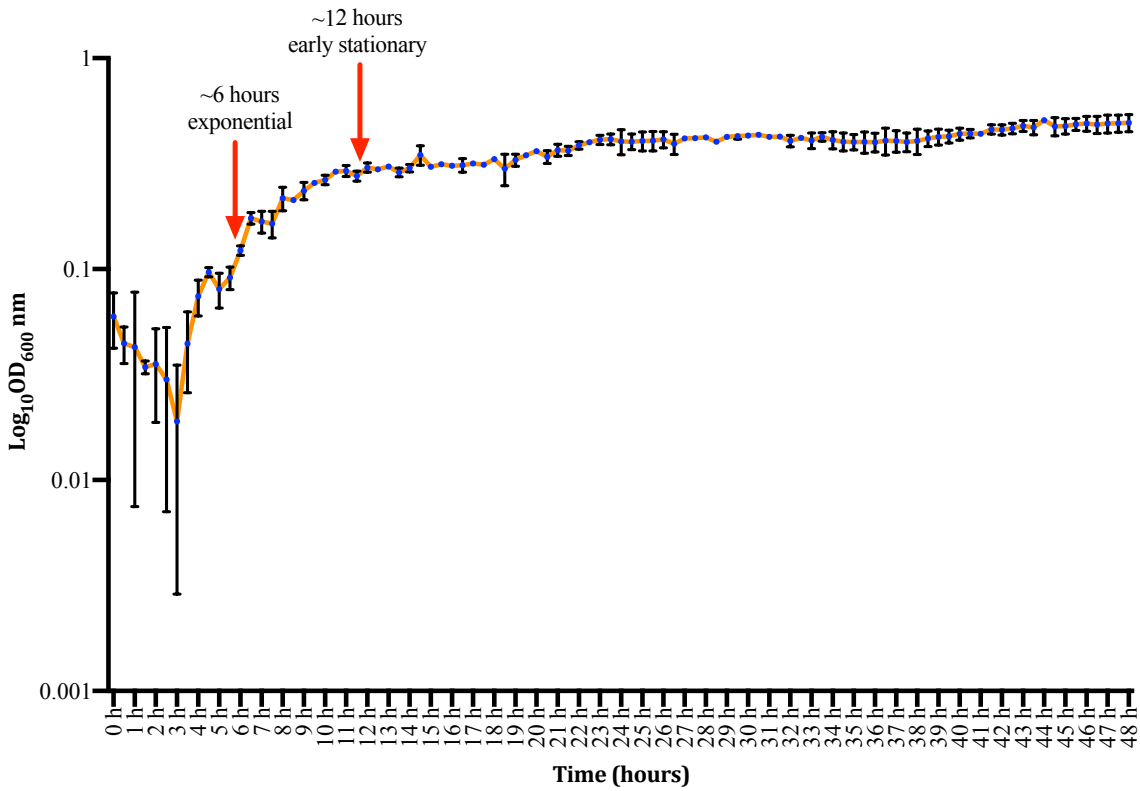


Figure 3.9: Log_{10} growth curve of NCTC 11168 in 0.1 % sodium deoxycholate. With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments. Red arrows indicate selected time points chosen for RNA extraction.

3.2.5 Iron limitation and repletion

Iron is a major virulence factor important for bacterial survival and when inside the human host, pathogenic bacteria are subjected to nutritional immunity with concentrations of essential metals such as restriction of iron in the host environment (Kortman *et al.*, 2014). To mimic an iron-limited environment, NCTC 11168 was grown in MEMα supplemented with 10 μM pyruvate. Growth was dramatically reduced without pyruvate supplementation and proved difficult to extract sufficient amounts of RNA from the low number of cells (data not shown). Iron repletion was achieved by adding 40 μM $FeSO_4$, showing improved growth, although not as significant when compared to nutrient rich MH2 broth (Figure 3.10). Iron-related growth conditions were adapted from Butcher and Stintzi (2013).

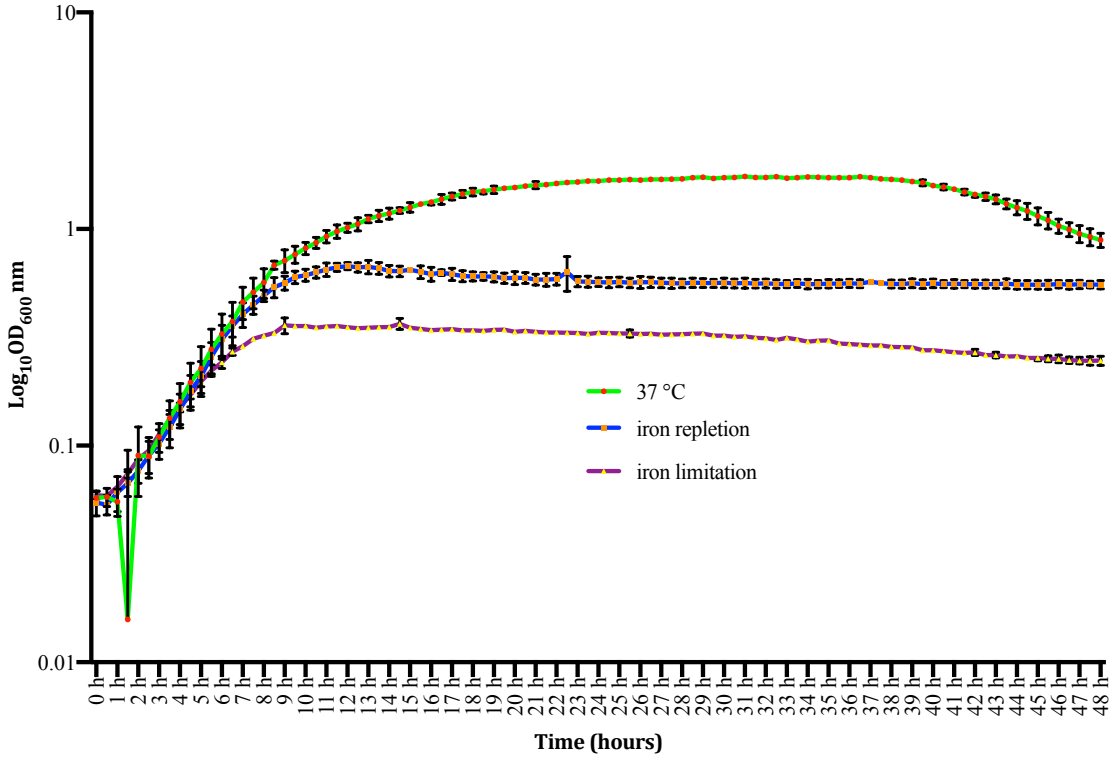


Figure 3.10: Different growth curves of NCTC 11168 under wild type and iron related conditions. The red and green line represents 37 °C MH2 broth, purple and yellow represents in MEMα supplemented with 10 μM pyruvate only, and orange blue represents MEMα supplemented with 10 μM pyruvate and 40 μM $FeSO_4$, with an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments.

Figure 3.11 shows the $\log_{10}$ growth curve of iron limitation vs iron repletion only. Both have the same pattern of growth but iron repletion reaches a higher OD_{600}. When harvesting cells for RNA extraction, iron limitation and repletion conditions were carried out in parallel on two separate days to collect a total of four overnight replicates. Exponential collection was taken at 4 hours 30 minutes OD_{600} 0.135-0.225 for both iron limitation and iron repletion. Early stationary collection was taken at ≈12 hours OD_{600} 0.225-0.27 for iron limitation and OD_{600} 0.72-0.765 for iron repletion.

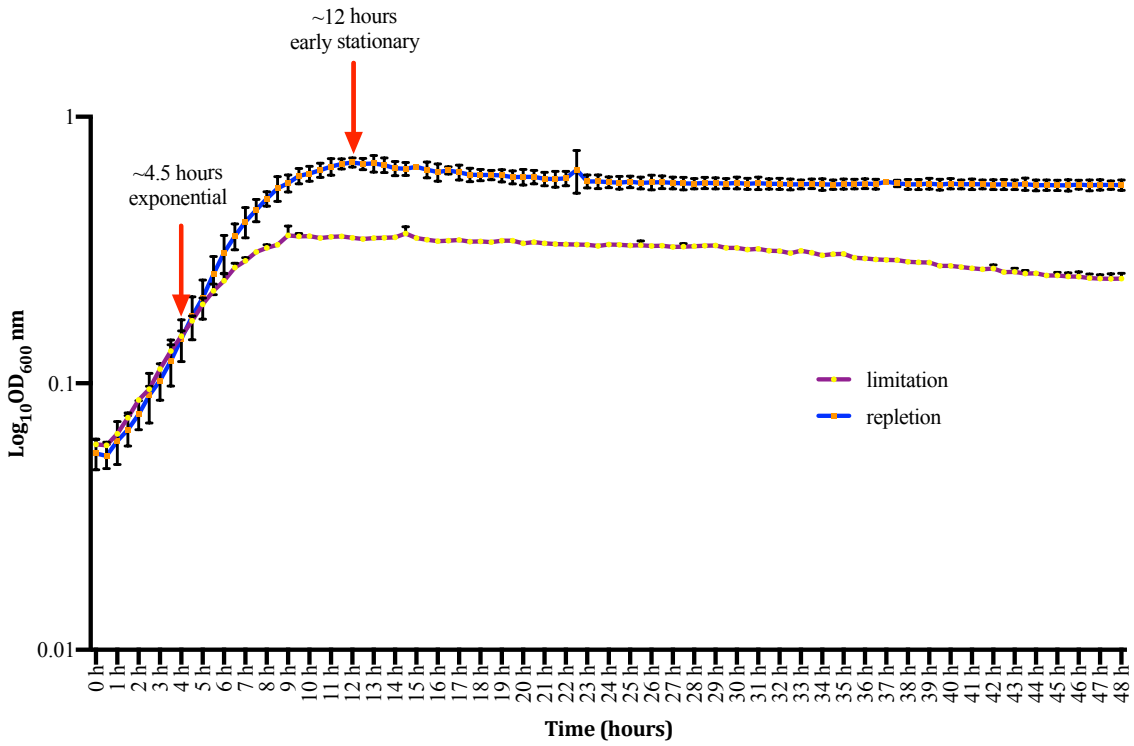


Figure 3.11: Log growth curves of NCTC 11168 at 37 °C in MEMα supplemented with 10 μM pyruvate only, and supplemented with 10 μM pyruvate and 40 μM $FeSO_4$. With an overall average of eight overnight technical replicates derived from two independent experiments that were carried out on different days. The standard deviation error bars were calculated from the two independent experiments. Red arrows indicate selected time points chosen for RNA extraction.

3.2.6 CFU counts

CFU counts of the cultures were taken before and after stresses were applied to ensure there was not more than a 1-fold log drop in numbers, and to see if the stress affected growth/viability. A minimum of three technical replicates (originating from separate agar plates) were used for statistical significance. All statistical tests for CFU counts were performed using a paired t-test if the data was determined to be normally distributed using the Shapiro-Wilk test, otherwise a wilcoxon signed rank test was used for calculating statistical significance.

3.2.6.1 Acid shock

Resuspending NCTC 11168 in MH2 broth adjusted to pH 3.5 for 10 minutes resulted in a significant drop in the CFU count with p value ≤ 0.0001 (Figure 3.12). This stress was adapted from Le *et al.* (2012).

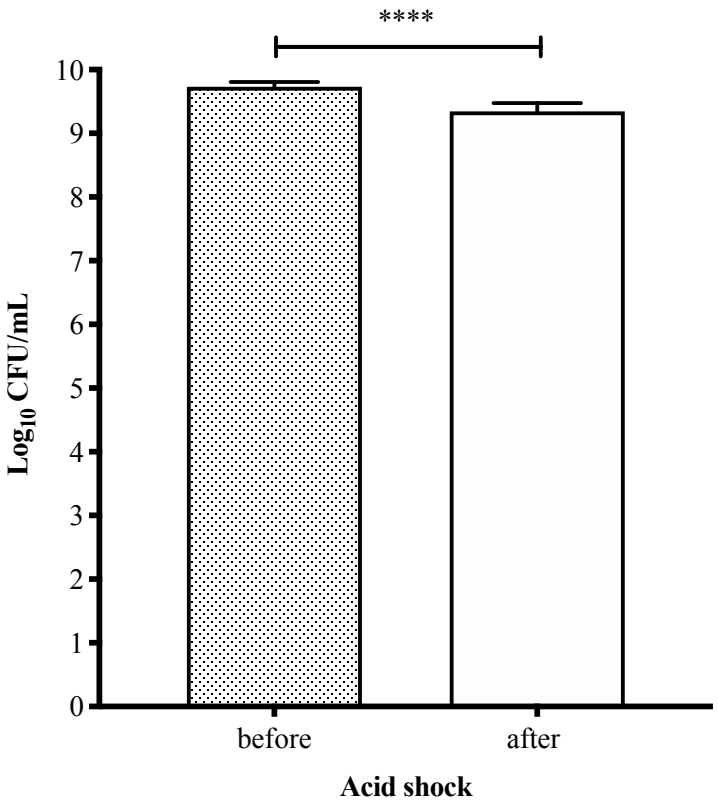


Figure 3.12: Bar graph of acid shock CFU/mL counts. With an average of nine replicates (three technical replicates plated thrice) before and after application of the stress and their significant difference. A paired t-test was used to calculate statistical significance and error bars show the standard deviation. **** = $p \leq 0.0001$

3.2.6.2 Heat shock

Heat shock was applied by transferring the 6-well plate to a 55 °C incubator in aerobic conditions for 3 minutes resulting in a significant drop in CFU with p value of 0.0381. This stress was adapted from Klančnik *et al.* (2014).

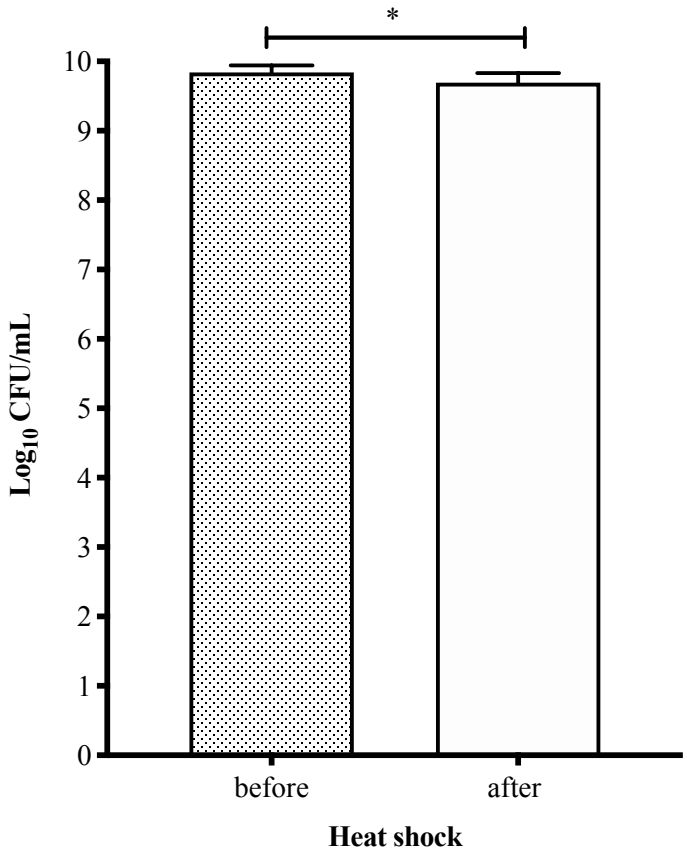


Figure 3.13: Bar graph of heat shock CFU/mL counts. With an average of 12 replicates (four technical replicates plated thrice) before and after application of the stress and their significant difference. A paired t-test was used to calculate statistical significance and error bars show the standard deviation. * = $p \leq 0.05$

3.2.6.3 Hyperosmotic stress

Osmotic stress was carried out for a total of 2 hours with two time points resuspended in 1.5 % NaCl. There was a significant drop in CFU after the first hour with a p value of 0.0002. However, there was a slight increase in growth in the second hour with a p value of 0.0321 between the first and second hour and 0.0246 between the original culture and the two hour stress (Figure 3.14). Two hours was chosen for RNA extraction. The stress was adapted from Cameron *et al.* (2012).

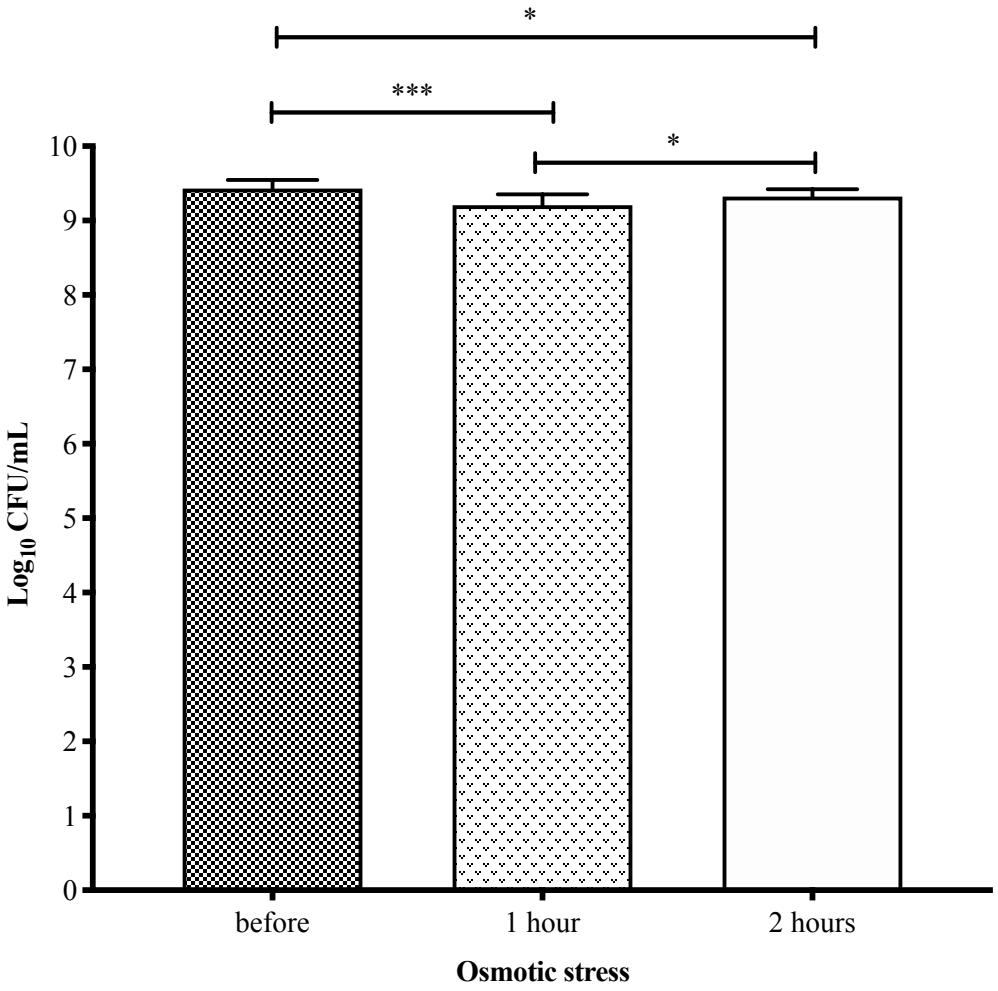


Figure 3.14: Bar graph of osmotic stress CFU/mL counts with an average of 12 replicates (four technical replicates plated thrice), before and after addition of 1.5 % NaCl for 1 and 2 hours and their significant difference. A paired t-test was used to calculate statistical significance and error bars show the standard deviation. * = $p \leq 0.05$ *** = $p \leq 0.001$

3.2.6.4 Anaerobic shock

Anaerobic shock was carried out by transferring the 6-well plate to an anaerobic chamber. There was no noticeable difference after 30 minutes but there was a slight increase in growth after 1 hour of incubation, although neither were significant (Figure 3.15). The latter was chosen for RNA extraction.

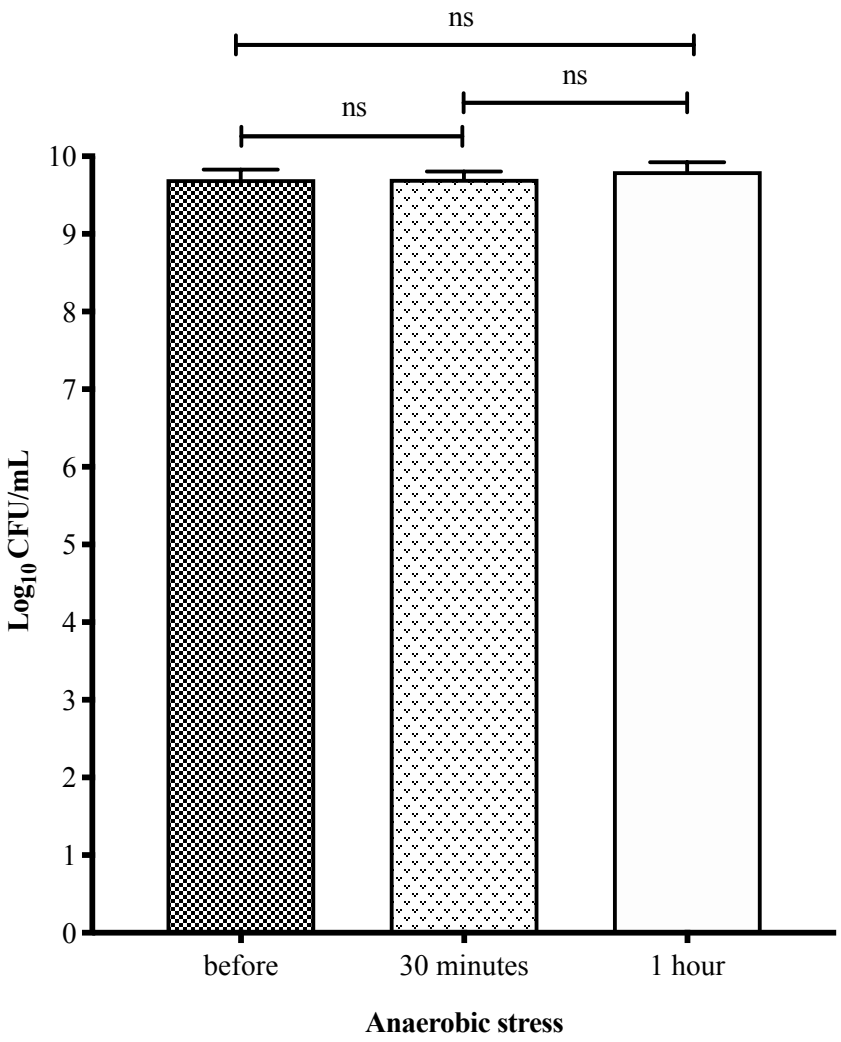


Figure 3.15: Bar graph of anaerobic CFU/mL counts. With an average of three biological replicates in technical triplicate before and after application of the stress for 30 minutes and 1 hour and their significant difference. A paired t-test was used to calculate statistical significance and error bars show the standard deviation. ns = no significance

3.2.6.5 Nutrient starvation

NCTC 11168 was grown to early stationary phase and resuspended in Ringer's solution for 5 hours. This stress had the biggest CFU drop with a p value of ≤ 0.0001 (Figure 3.16). This stress was adapted from Klančnik *et al.* (2009).

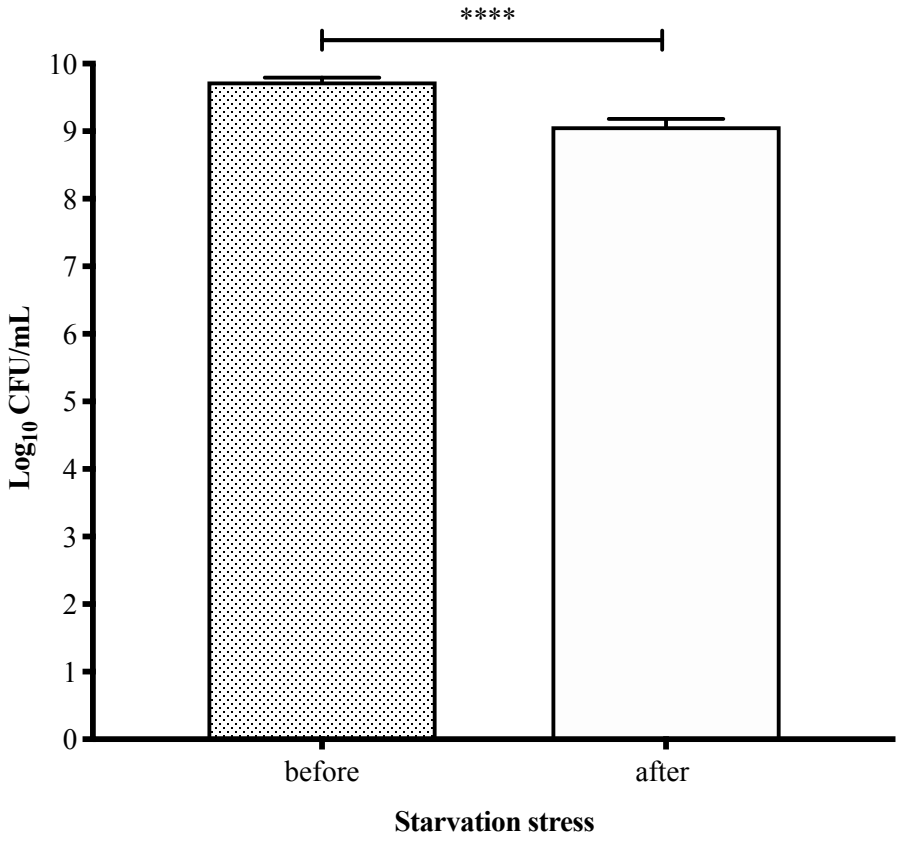


Figure 3.16: Bar graph of nutrient starvation CFU/mL counts. With an average of 12 replicates (four technical replicates plated thrice), before and after application of the stress and their significant difference. A paired t-test was used to calculate statistical significance and error bars show the standard deviation. **** = $p \leq 0.0001$

3.2.6.6 Peroxide stress

3 mM of hydrogen peroxide was added to NCTC 11168 cultures for 10 minutes and incubated in microaerophilic conditions. There was a significant drop in CFU counts with a p value of ≤ 0.001 (Figure 3.17). This stress was adapted from Klančnik *et al.* (2006).

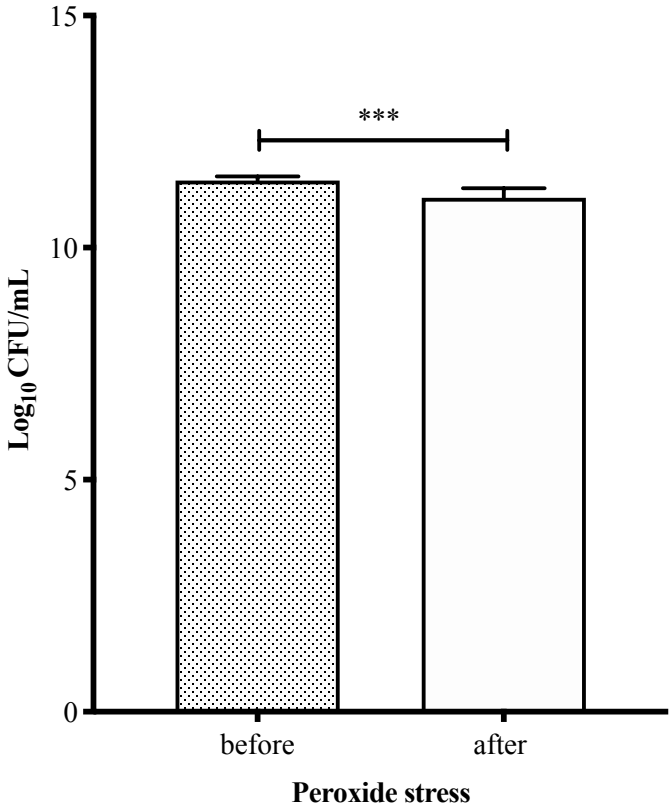


Figure 3.17: Bar graph of peroxide stress CFU/mL counts. With an average of 12 replicates (four technical replicates plated thrice), before and after application of the stress and their significant difference. The wilcoxon signed rank test was used to calculate statistical significance and error bars show the standard deviation. *** = $p \leq 0.001$

3.2.6.7 Nitrosative stress

Different concentrations of GSNO: 200 μM, 400 μM, and 600 μM of GSNO were added to NCTC 11168 but had no noticeable difference in CFU counts (data not shown). The maximum concentration without being toxic to cells used in Elvers *et al.* (2005) was 1.5 mM GSNO and this was the final concentration used for applying nitrosative stress. 1.5 mM GSNO was added to NCTC 11168 cultures for 1 hour and incubated in microaerophilic conditions shaking at 200 rpm. There was no significant change in CFU counts (Figure 3.18). This stress was adapted from Elvers *et al.* (2005).

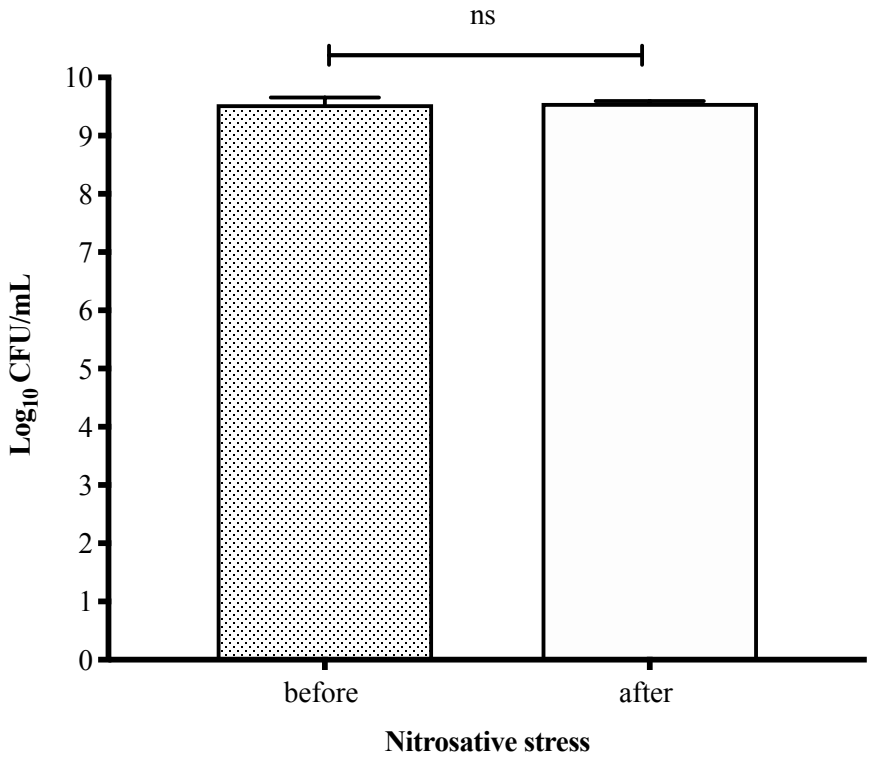


Figure 3.18: Bar graph of nitrosative stress CFU/mL counts. With an average of three biological replicates in technical triplicate, before and after addition of 1.5 mM GSNO and their significant difference. The wilcoxon signed rank test was used to calculate statistical significance and error bars show the standard deviation. ns = no significance

3.2.6.8 Chicken juice and cold stress

NCTC 11168 was resuspended in either 100 % chicken juice, diluted chicken juice (5 % v/v) in MH2 broth, or MH2 broth only and incubated at 4 °C in an AnaeroJar in microaerophilic conditions. There was no significant difference in CFU counts before and after resuspension in 100 % chicken juice, but significant drop in CFU counts for MH2 broth supplemented with 5 % chicken juice and MH2 broth only with p values of ≤ 0.0001 and $p = 0.0002$ respectively. The stresses were adapted from Brown *et al.* (2014).

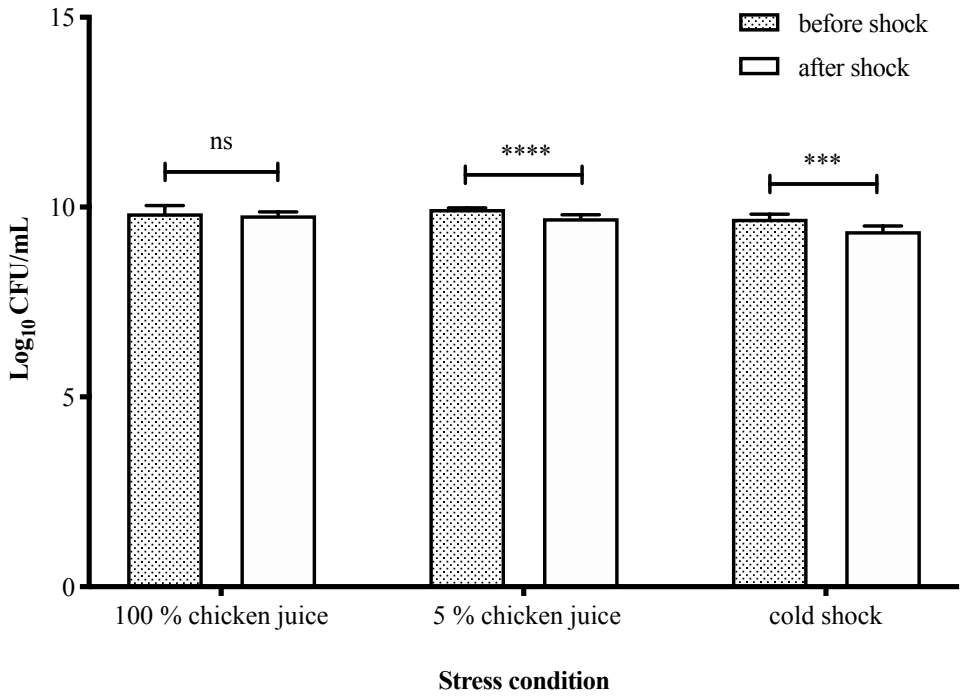


Figure 3.19: Bar graph of 100 % chicken juice, MH2 supplemented with 5 % chicken juice, and MH2 broth only (cold shock) control CFU/mL counts. With an average of 12 replicates (four technical replicates plated thrice) before and after viability test and their significant differences. A paired t-test was used for 5 % chicken juice and the wilcoxon signed rank test was used for 100 % chicken juice and cold stres to calculate statistical significance. Error bars show the standard deviation. ns = no significance, **** = $p \leq 0.0001$, *** = $p \leq 0.001$

3.3 Optimisation of RNA extraction methods

Three RNA extraction methods: hot-phenol, bead-beating, and miRNAeasy Mini Kit (Qiagen) were chosen for total RNA extraction to test for the quality of RNA and yield and sequenced on the Illumina MiSeq. The methods are described in detail in Section 2.3.

To compare RNA extraction methods, NCTC 11168 was grown in MH (Oxoid) and 1 mL of culture was collected at exponential phase at an OD_{600} of ≈0.45-0.54, then the remaining 2.5 mL was grown to early stationary phase OD_{600} at 1.3-1.8 (Figure 3.1). Cells were grown from the same overnight cultures in order to compare each method with each other. Exponential and early stationary cells were also taken from the same 6-well plate (Greiner) and pooled together before undergoing extraction. The growth and harvesting of cells was repeated on a different day to represent two biological replicates.

Each technical replicate per biological replicate was subjected to either hot-phenol, bead-beating, or miRNAeasy Mini kit (Qiagen) RNA extraction (detailed in section 2.3). RNA was DNase treated according to each protocol and quality checked. The concentration was measured using Qubit® 2.0 Fluorometer (Invitrogen) shown in Table 3.1, each method did not vary greatly although miRNeasy Mini kit had a slightly lower yield.

Extraction method		Amount of RNA (ng)
hot-phenol	A	11,400
	B	25,700
bead-beating	A	13,360
	B	22,800
miRNAeasy Mini kit (Qiagen)	A	14,220
	B	8,040

Table 3.1: Comparison of total RNA (ng) between different RNA extraction methods. This was measured post DNase treatment using the Qubit RNA-Broad range kit. A and B indicate the two individual experiments carried out.

There was also not much difference between RNA quality as seen in Figure 3.20. The Bioanalyzer trace shows that all samples had intact 16S and 23S ribosomal RNA (rRNA) peaks. All samples had a RNA integrity value (RIN) value of 10 apart from hot-phenol replicate A which had a RIN value of 9.8. A RIN value of 8 or above is adequate for RNA-seq.

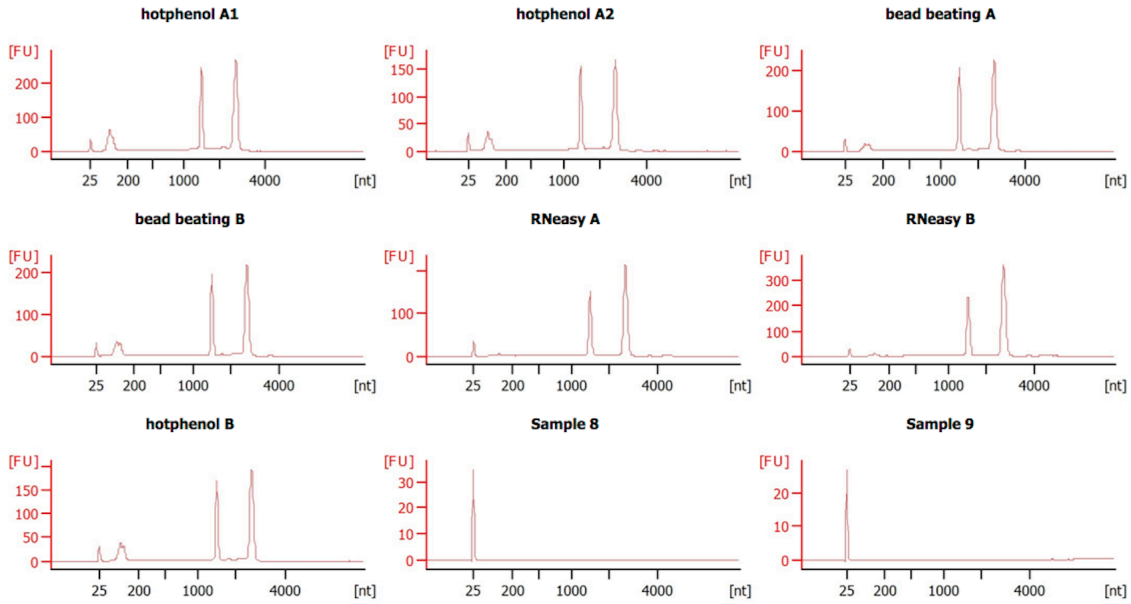


Figure 3.20: Bioanalyzer traces of RNA extractions post DNase treatment. The first peak shows the 5S ribosomal RNA (rRNA), second peak 16S rRNA, and third peak 23S rRNA. There is a marker at the beginning $\approx$ 25 nt as seen in the empty sample wells 8 and 9. N.B. Hot-phenol replicate A was split into two samples A1, and A2 during DNase treatment.

After ribosomal depletion using Ribo-Zero^{TM}, cDNA libraries were made using the TruSeq stranded Total RNA library preparation kit (Illumina), quality checked and ran on a Illumina MiSeq v2 cartridge 2 x 76 cycle paired-end sequencing (one extra cycle for phasing). For normalization, 3,000,000 reads were randomly sub-sampled from each fastq file as this was the lowest number of reads from the 6 samples using the following command:

seqtk sample -s100 [fastq file] 3000000 > [output file]

Fastq files were converted to sorted BAM files and viewed on Artemis: sequence visualization and annotation (Carver *et al.*, 2012).

3.3.1 Identification of ncRNAs between different RNA extraction methods

To examine which RNA extraction method was able to retain the most ncRNAs, toRNAdo[1] was used for identification of putative ncRNAs. The tool generated wig files and an output csv file containing information on putative intergenic, antisense, or mixed non-coding regions. These included tRNAs and unannotated gene features such as the selW orthologue (Gursinsky *et al.*, 2008). Bead-beating found the most putative ncRNAs in both biological replicates after sub-sampling compared with hot-phenol and miRNeasy mini kit (Figure 3.21). Surprisingly, hot-phenol had the least number of ncRNAs found and only had one unique ncRNA in both biological replicates with little overlap between the other methods. There were ncRNAs found in all three methods but also many unique to bead-beating and the miRNeasy kit.

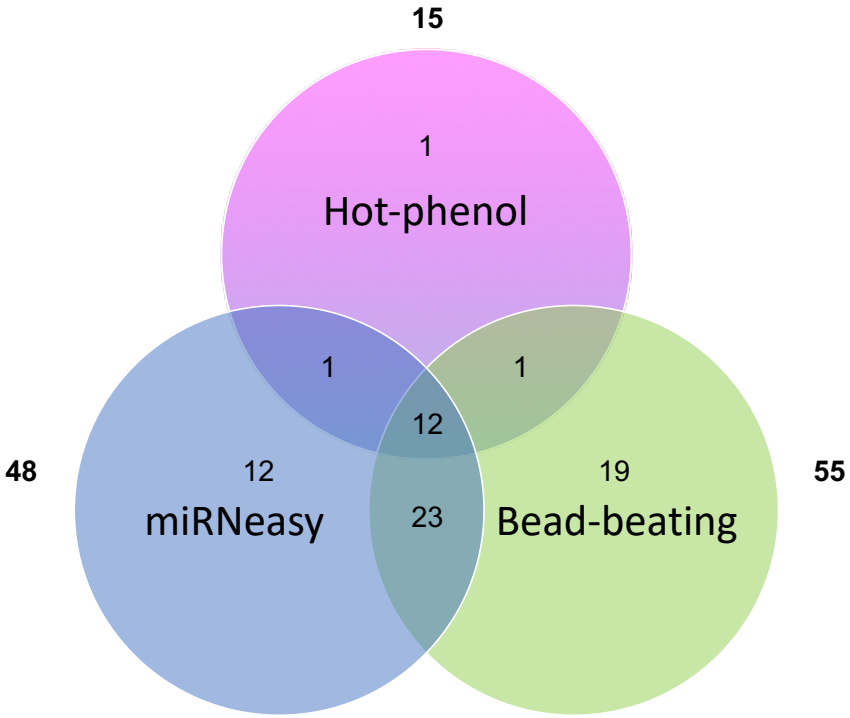


Figure 3.21: Venn Diagram of total putative ncRNAs in three different RNA extraction methods: Hot-phenol, bead-beating, and miRNeasy mini kit (miRNeasy) in both biological replicates after sub-sampling including un-annotated gene features and tRNAs. Numbers in bold depict the total number of putative ncRNAs that were identified for each method.

[1] `https://github.com/pavsaz/toRNAdo`

3.4 Discussion

3.4.1 Optimisation of growth curves and stress conditions

Optimisation of NCTC 11168 growth curves in a 6-well plate (Greiner) carried out in a FLUOStar Omega microplate reader (BMG labtech) were successful in generating reproducible growth curves. This was true even for growth conditions using different media/stresses such as MH broth and MEMα. Our growth curves also corresponded well with published characterised growth phases in a VAIN at 42 °C: exponential phase (0-16h), retardation (16h-20h), stationary phase (20-40h), and decline phase (40-66h) (Wright *et al.*, 2009). One downside of this method is it is not possible to take CFU counts during growth as this would disturb environmental conditions and reduce the final volume of the culture. Even so, CFU is not an accurate determinant for the number of cells due to clumping and the VBNC state occuring in the later growth phases. It was also shown that different brands can affect growth as growth curves using MH (Oxoid) and MH2 (Sigma-Aldrich) did not show the same trend (Figure 3.4).

Stress conditions were optimised by consulting the literature and ensuring the stress applied was mild enough that it did not elicit a general stress response and cause major cell death but was able to induce transcriptional changes presumable from the applied stress. Although *C. jejuni* does evoke a stringent response, it is atypical compared to other Gram-negative bacteria as it is primarily regulated by SpoT rather than the classic RpoS regulator (Gaynor *et al.*, 2005). CFU counts were taken before and after the stress was applied to observe the effect on cell numbers, and to ensure there was not more than 1 log drop in CFU. As mentioned previously, clumping and the VBNC state may be issues, but the focus is only on the relative comparison before and after the stress and the majority of stress conditions were applied at exponential phase. Although not a comprehensive list, the growth and stress conditions are representative of the lifestyle *C. jejuni* encounters during human infection, chicken colonisation, and transmission. These conditions were also chosen for their ease of replication in the laboratory and availability of published methods to reduce the time taken for optimisation. Some of these conditions have had their transcriptional profile studied before, which can be used as a reference. The standard routine laboratory conditions were included and 37 °C exponential phase was selected as the control group for many comparisons with the other stress conditions.

It was observed that *C. jejuni* grown at 42 °C grows slighter faster than at 37 °C (Figure 3.7). This contradicts Khanna *et al.* (2006)'s study where growth rate was greater at 37 °C compared to 42 °C. However, there were differences in the experimental setup by Khanna *et al.* (2006) such as the use of Kapadnis-Baseri (KB) broth rather than MH2 broth, no shaking, and the *C. jejuni* strain was an unknown isolate. Either way, this illustrates that a 5 °C temperature difference already has a slight effect on growth and may be important for deciphering the differential behaviour of *C. jejuni* between chicken commensal and human host.

Sodium deoxycholate stress and iron limitation and repletion conditions showed a significant drop in growth rates compared to standard growth conditions. Pellets recovered from these stresses had lower RNA yield due to repressed growth. In particular, pellets from early stationary sodium deoxycholate stress had a loss of typical pink pigmentation found in *C. jejuni*. A similar observation was seen in Liu and Kelly (2015)'s study where pellets of the cccA (Cj1153) mutant (involved in cytochrome biogenesis) were brown rather than pink. The decrease in growth with sodium deoxycholate stress also correlated with Negretti *et al.* (2017)'s study where OD_{540} dropped after ≈12 hours for three strains: 81-176, F38011, and NCTC 1168 with MH broth supplemented with 0.05 % (w/v) sodium deoxycholate compared to absence of sodium deoxycholate, and there was also a decrease in CFU counts. In this project, 0.1 % sodium deoxycholate was supplemented in MH2 broth and there was a decrease in growth when compared to the absence of sodium deoxycholate from the beginning with the biggest difference observed around 10 hours as seen in Figure 3.8. Bile salts are known to have antimicrobial effects and a study by Cremers *et al.* (2014) showed that the bile salts cholic acid and deoxycholic acid (i.e. sodium deoxycholate) are protein-unfolding reagents in *E. coli*. Negretti *et al.* (2017) also speculates that oxidative damage of DNA occurs and membrane permeability is altered in the presence of bile. It is unclear whether sodium deoxycholate is a chemorepellent or chemoattractant from Li *et al.* (2014)'s study, but it is clear from this project that it is unfavourable for growth.

It is interesting to note that *C. jejuni* was more viable in 100 % chicken juice at 4 °C than in MH2 supplemented with 5 % chicken juice or MH2 broth only (Figure 3.19). This suggests there are some components in 100 % chicken juice which gives protection or provides nutrients to *C. jejuni* to allow it to stay viable for long periods at 4 °C. A study by Birk *et al.* (2004) also found that *C. jejuni* remained viable longer in 100 % chicken juice than Brain Heart Infusion (BHI) media supplemented with 5 % calf blood with only a slight decrease in CFU after 5 days of incubation at 5 °C. Chicken juice was also shown to increase biofilm formation and enhance *C. jejuni* growth (Brown *et al.*, 2014). This could be a reason why 100 % chicken juice at 4 °C did not have a huge drop in CFU counts, unlike 5 % chicken juice or only MH2 broth. However, *C. jejuni* biofilm production was markedly decreased in nutrient-rich media (Reeser *et al.*, 2007). Biofilm enhancement in the presence of chicken juice is due to providing a conditional layer for biofilms to form on.

The biggest drop in CFU counts was during starvation stress which was applied at early stationary phase, though this could be a result of entering the VBNC state and changing morphology to become coccoid shaped as was shown in Klančnik *et al.* (2009)'s study observed under the microscope. Some stresses, however, only showed a very slight drop in numbers or even grew in numbers as seen in anaerobic stress and nitrosative stress, although the growth was not significant. This could be an indication that bacteria are adapting to stress rather than fighting for survival. For example, the *C. jejuni* strain CI 120 elicits an adaptive tolerance response that is exclusive to this strain at early stationary phase to mild acidic and aerobic conditions, with increased acid resistance post-exposure (Varsaki *et al.*, 2015). At high temperatures, a 3 minute incubation of 55 °C was enough to cause a significant drop in CFU counts (Figure 3.13) since *C. jejuni* cell death starts occurring at 56-57 °C (Nguyen *et al.*, 2016). This was incubated under aerobic

conditions as the plate reader only has a range of 25-45 °C, but it is speculated that the atmosphere in the 6-well plate without removing the lid does not change oxygen levels drastically and rapidly within 3 minutes to calibrate to aerobic conditions. Acid and peroxide stress also had a significant drop in CFU with only 10 minutes exposure to the stress, suggesting the conditions applied may be detrimental to *C. jejuni*. NCTC 11168 is not known to have an adaptive tolerance response.

3.4.2 Optimisation of RNA extraction methods

In the literature, hot-phenol RNA extraction method is considered the best method for extracting high quality RNA and was the method of choice used in similar studies in *C. jejuni* (Butcher and Stintzi, 2013; Dugar *et al.*, 2013; Porcelli *et al.*, 2013). A study by Jahn *et al.* (2008) also found hot-phenol to be the optimal RNA extraction method for a plant pathogen *Dickeya dadantii* when compared with stand alone and hybrid combinations of Trizol (Invitrogen) and RNeasy (Qiagen) kits (Jahn *et al.*, 2008). Hot-phenol, however, is very dangerous to handle and not an easy protocol to implement. Commercial kits are available which offer an easier and safer method, although they can be very expensive and normally use columns, which may not retain small sizes of RNA. There was not much difference found in the yield or RIN values between the different methods. All the methods are phenol-based but the major difference is the lysing step. Bead-beating has the most rigorous lysing treatment and seems to lyse the cells better compared to the other methods and therefore gives a more accurate representation of RNAs that were transcribed in cells. Lysing was also shown to correlate with good RNA yield in Jahn *et al.* (2008)'s study. As bead-beating identified the most ncRNAs in both replicates compared with hot-phenol and miRNeasy mini kit, all subsequent RNA extractions throughout this project utilised this method.

The bead-beating protocol was further optimised over the duration of the project (see Section 2.3.4). Acidic phenol pH 4.5 125:24:1 phenol:chloroform:isomyl alcohol was used instead of 25:24:1 phenol:chloroform:isomyl alcohol to recover less DNA as a 5:1 ratio should result in absence of DNA in the upper aqueous phase (Zumbo, 1979). Other changes include using isopropanol for overnight precipitation rather than using the RNA clean and concentrator^{TM} -25 kit to save costs and prevent any loss of RNA through repeated cleaning runs. Incubating the RNA for 3 hours on ice ensures the RNA pellet is properly dissolved before snap-freezing in liquid nitrogen for long term storage. There was distinctly less DNA contamination using this method compared to all the previous methods.

3.4.3 Summary

In conclusion, optimisation of methods for this project was carried out successfully with defined growth and stress conditions. MH2 cation-adjusted broth was found to be a more reproducible and standardised media for culturing NCTC 11168 and should be considered the standard of use when using MH broth for *C. jejuni* maintenance and growth. In general, MH is already widely used for testing antimicrobial compounds and yields the best growth for *C. jejuni* (Davis and DiRita, 2008; Ng *et al.*, 1985). CFU counts demonstrated the stresses that were applied were representative of the initial exposure to the stress that either elicit a transient transcriptional response for adaptation or long-term stress for survival. Established RNA extraction methodologies in the literature were not optimised for the recovery of ncRNAs. Our data suggests that bead-beating was the optimal RNA extraction method for the purposes of identifying ncRNAs. Nevertheless, the other two RNA extraction methods that were compared in this chapter, in particular the miRNeasy kit (Qiagen), is adequate for normal RNA-seq and depending on the research question asked, if ncRNAs are not of interest then alternative and easier methods are certainly more applicable. These optimised methods were used throughout this project to produce a good set of samples to obtain data for Chapters 4 and 5.

Chapter 4

Identification of TSS under a pool of 22 host-relevant stress conditions

4.1 Introduction

The primary transcriptome is represented by the entirety of primary transcripts derived from the total RNA in a population of cells used for genome-wide identification and mapping of transcriptional start sites (TSS). A TSS is the first nucleotide location from where transcription of a RNA molecule starts at the 5' end. Detecting and knowing where TSS are placed uncovers promoter sequences that give insights into: DNA binding motifs, sigma factor preferences, multiple TSS, and overall transcriptional organisation (Schlüter *et al.*, 2013; Kapranov, 2009). As no prior information of the sequence is needed for detection of TSS using high throughput sequencing (HTS) technologies, novel transcripts can also be found this way.

It is important to distinguish between primary transcripts that have 5' triphosphate (5'-PPP) ends and processed transcripts that have 5' monophosphate (5'-P) ends, as the latter is typically derived from degradation or processing of RNA into rRNA or tRNA (Kapranov, 2009). Current methods to determine the primary transcriptome rely on this distinction. One of the earlier methods to differentiate between these transcripts is differential RNA-seq (dRNA-seq) by Sharma *et al.* (2010). This method uses the enzyme terminator exonuclease (TEX) to preferentially degrade 5'-P ends, so it essentially depletes processed transcripts thereby enriching for primary transcripts (Sharma and Vogel, 2014). Since then, many other methods to determine the primary transcriptome have emerged such as tagRNA-seq, 5'end-Seq, and Cappable-seq (Innocenti *et al.*, 2015a; Wurtzel *et al.*, 2012; Ettwiller *et al.*, 2016).

A caveat of using dRNA-seq is that TEX may not efficiently degrade all processed transcripts that have secondary structures as they obstruct the activity of the enzyme leading to imprecise detection of TSS (Ettwiller *et al.*, 2016). Many of the alternative methods use the same principle as dRNA-seq but use different approaches to distinguish primary and processed transcripts. TagRNA-seq differentiates primary and processed transcripts by initially tagging processed 5'-P ends, before using tobacco alkaline

phosphatase (TAP) to degrade 5'-PPP to 5'-P ends, and then uses a different tag for the remaining treated primary transcripts (Innocenti *et al.*, 2015a). Advantages using this method is that information on processed transcripts is not lost, and it is possible to quantify an estimate of spurious transcripts. However, although the majority of TSS were detected and mapped at a single-nucleotide resolution, applications of the methodology state there were cases where accuracy of mapping was at 6 nt or less (Innocenti *et al.*, 2015a,b). 5'end-Seq uses a similar concept to dRNA-seq but opts for degrading the primary transcripts using TAP for apparent enrichment of 5'-PPP (Wurtzel *et al.*, 2012). For Cappable-seq, 5'-PPP ends are capped with 3' desthiobiotin (DTB-GTP) and captured using streptavidin beads, which have an affinity for biotin, so rather than degrading the processed transcripts for an apparent enrichment there is direct enrichment of the 5'-PPP primary transcripts (see Figure 2.2 in Section 2.5.3) (Ettwiller *et al.*, 2016). Cappable-seq treated samples of *E. coli* were shown to deplete rRNA and the identified TSS have good overlap when compared with dRNA-seq data in the same organism as well as finding many additional putative novel TSS (Ettwiller *et al.*, 2016). This method was chosen to determine the primary transcriptome of NCTC 11168 at 1 bp resolution under 22 conditions, as there is no reliance on degradation of RNA transcripts.

RNA-seq data is overwhelmed by background noise primarily due to saturated signals from rRNA and tRNA. Indeed, rRNA makes up ≈95 % of the total RNA in prokaryotes (Giannoukos *et al.*, 2016; Ettwiller *et al.*, 2016). In order to reduce this, rRNA depletion kits such as Ribo-Zero^{TM} (Illumina) are commonly used. Ribo-Zero^{TM} was shown to be the most efficient at depleting rRNA and targets 5S rRNA unlike other kits, although it is not effective against 5S rRNA in *C. jejuni* (Petrova *et al.*, 2017). It has also been reported that rRNA depletion introduces a sequencing coverage bias (Lahens *et al.*, 2014). Not depleting rRNA would require deeper sequencing as rRNA would otherwise trump any meaningful signal, especially since some species have several copies of rRNA such as three copies in *C. jejuni* (Espejo and Plaza, 2018). Cappable-seq, however, does not require rRNA depletion of the initial total RNA sample, as it was shown that the enrichment step does not capture rRNA and therefore depletes it (Ettwiller *et al.*, 2016). Depletion of rRNA can therefore be used as an indication and a measure of successful enrichment.

Determining TSS can allow for novel transcripts to be discovered such as ncRNAs. To date there are two other published studies on the primary transcriptome of *C. jejuni* NCTC 11168 in the literature by Dugar *et al.* (2013) and Porcelli *et al.* (2013), although these were only carried out in standard laboratory conditions, and one study that determined TSS of PerR-regulated and Fur-repressed genes from a *fur perR* mutant by Handley *et al.* (2015). Many promoters for annotated genes are still yet to be discovered as they are likely induced by certain environmental conditions or stresses, and the functions of many genes in *C. jejuni* remain a mystery. Moreover, these studies identified many novel putative ncRNAs, but many may have been missed as ncRNAs are mainly induced by stress (Hoe *et al.*, 2013). By using a pooled data set of host-relevant and transmission conditions, it is anticipated that many more putative novel transcripts and promoters will be identified that were previously missed in published studies. Indeed, primary transcriptomes carried out in a compendium of conditions in other bacteria have identified

novel ncRNAs in their genomes such as for *S. enterica* Typhimurium, *S. pneumoniae*, and *Burkholderia pseudomallei* (Kröger *et al.*, 2013; Ooi *et al.*, 2013; Slager *et al.*, 2018). This gives a global overview of the transcriptomic landscape of *C. jejuni* NCTC 11168 under a wide range of stresses providing a useful resource for the *Campylobacter* research community.

To assess whether the data obtained from Cappable-seq is robust and useful for revealing novel architecture, iron homeostasis was investigated in more detail as it is a well-studied aspect of *C. jejuni* microbiology with available published microarray and transcriptomic data (Palyada *et al.*, 2004; Holmes *et al.*, 2005; Butcher *et al.*, 2012, 2015). The primary transcriptome has yet to be determined under iron-limited or repleted conditions, so we envisage to find novel condition-dependent promoters for iron-regulated genes that previously had no annotated promoters.

Chapter Aims:

- Determine the primary transcriptome from a pool of 22 conditions.
- Compare TSS candidates with published data and assess the reproducibility of Cappable-seq.
- Identify novel transcripts that are condition-dependent and look for promoter motifs.
- Evaluate the data set as a potential resource for the *Campylobacter* community.

4.2 Workflow and quality control of Cappable-seq data

The same RNA samples extracted from 22 different growth and stress conditions as discussed in Chapter 3 and analysed in Chapter 5 were pooled together to generate Cappable-seq libraries (the full protocol can be accessed in Chapter 2, Section 2.5.3). The pooled RNA was enriched for primary transcripts using the Cappable-seq system with a control sample and made into a specific sRNA library to determine TSS. After sequencing on the Illumina MiSeq platform, FastQC was carried out to check the quality of the fastq files before and after trimming. Although there were some over represented sequences, the quality scores for both the control and treated samples were very good. Bowtie2 was used to generate an indexed reference genome from a fasta file of strain NCTC 11168 taken from NCBI Accession number: AL111168 AL139074-AL139079. The indexed genome was subsequently used to map reads from the fastq files. The alignment rate for NCTC 11168 was very high as seen in Table 4.1. Only the mapped reads were used for further analysis.

Sample	Total number of reads	Overall alignment rate
Control	16,264,529	93.26 %
Enriched	11,016,457	96.19 %

Table 4.1: Table of the total number of reads and percentage alignment of Cappable-seq samples. Only the mapped reads were used for subsequent analysis.

Bam files of both control and treated samples were visually inspected on Artemis (Carver *et al.*, 2012). The raw reads mapping to rRNA were extracted from the bam files on Artemis to calculate percentage of reads mapping to rRNA regions. The treated enriched sample had 1.86 % reads whereas the non-treated control had 73.63 % reads mapping to rRNA indicating that enrichment using Cappable-seq was successful. Using the Perl script bam2firstbasegtf.pl (see Chapter 2, Section 2.7.3 for methodology), all the mapped reads were trimmed to their individual first nucleotide corresponding to the most 5' nucleotide of each primary transcript mapping to the genome, thus enabling identification of TSS at 1 bp resolution. To account for differences in sequencing depth between the two samples, the number of reads mapping to each TSS was normalised using the relative read score (RRS): the number of reads mapping at each TSS position and orientation (+/-) of strand divided by the total number of mapped reads multiplied by 1,000,000. A cut-off of 1.5 RRS or higher was used for identifying TSS in the enriched bam file, whereas no cut-off was used for the control sample.

The treated enriched bam file was also trimmed to 1 bp for visualisation on Artemis as seen in Figures 4.1 to 4.3. In addition, the published annotated Artemis file containing promoters from Porcelli *et al.* (2013) was viewed on Artemis for a quick indication of whether any identified Cappable-seq TSS are novel. The peaks shown correspond to where the reads map to with: a published annotated promoter, no published annotated promoter present (a novel TSS), and an annotated internal TSS with an antisense RNA (asRNA)

TSS on the opposite strand for (Figures 4.1 to 4.3) respectively. The modified enriched bam file however only contains the raw mapped reads without normalisation or any further processing.

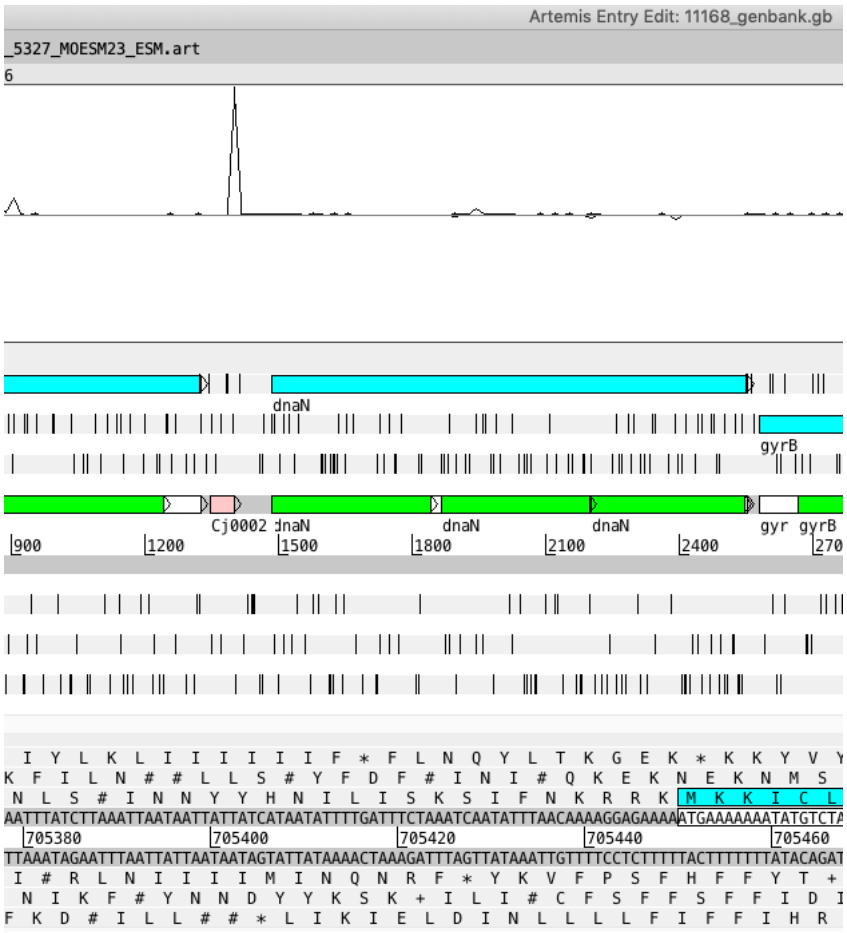


Figure 4.1: A screenshot of the coverage from the enriched raw bam file that corresponds to a published annotated promoter viewed on the Artemis genome browser. NCTC 11168 GenBank file and an annotated Artemis file with promoter information by Porcelli *et al.* (2013) was also loaded onto Artemis. The screenshot shows an expression peak for the annotated promoter Cj0002 denoted by the pink box upstream of the gene *dnaN*. The TSS is at the 3' end of the annotated pink promoter.

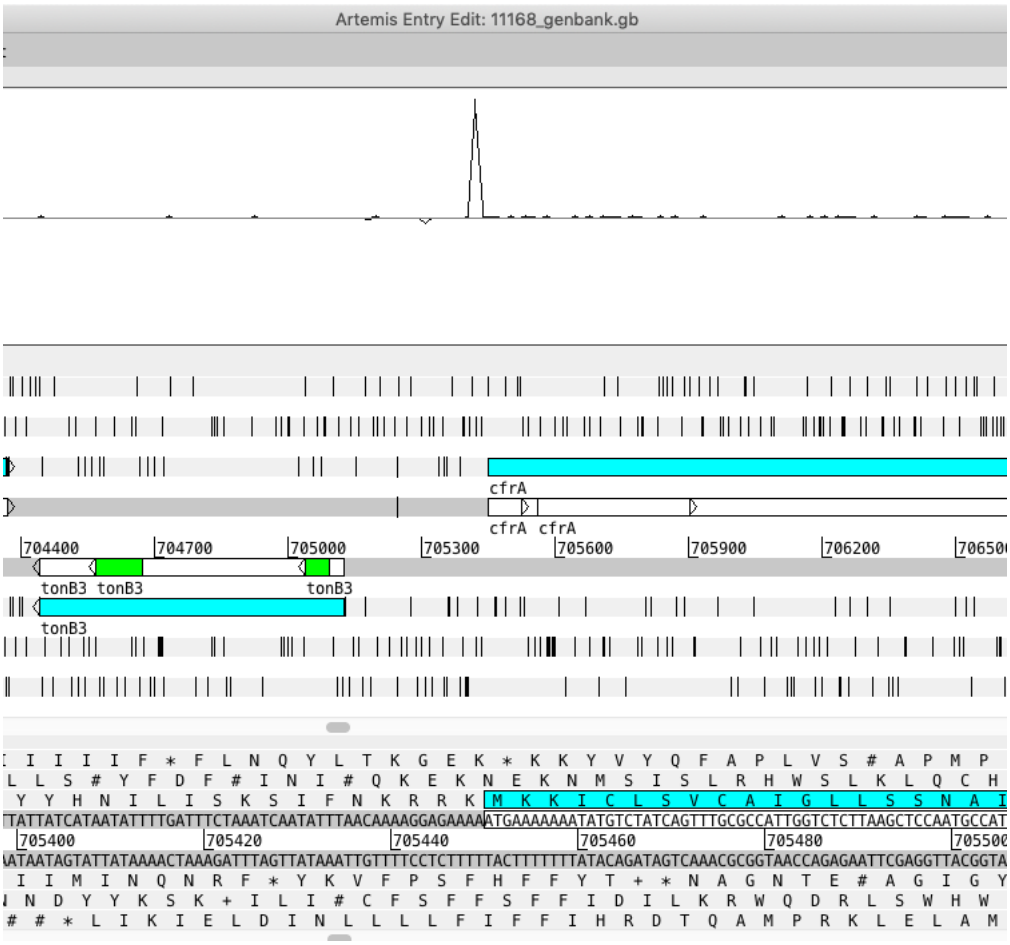


Figure 4.2: A screenshot of the coverage from the eniched raw bam file that corresponds to a potential novel TSS viewed on the Artemis genome browser. NCTC 11168 GenBank file and an annotated Artemis file with promoter information by Porcelli *et al.* (2013) was also loaded onto Artemis. The screenshot shows an expression peak upstream of the gene *cfrA* with no annotated promoter present.

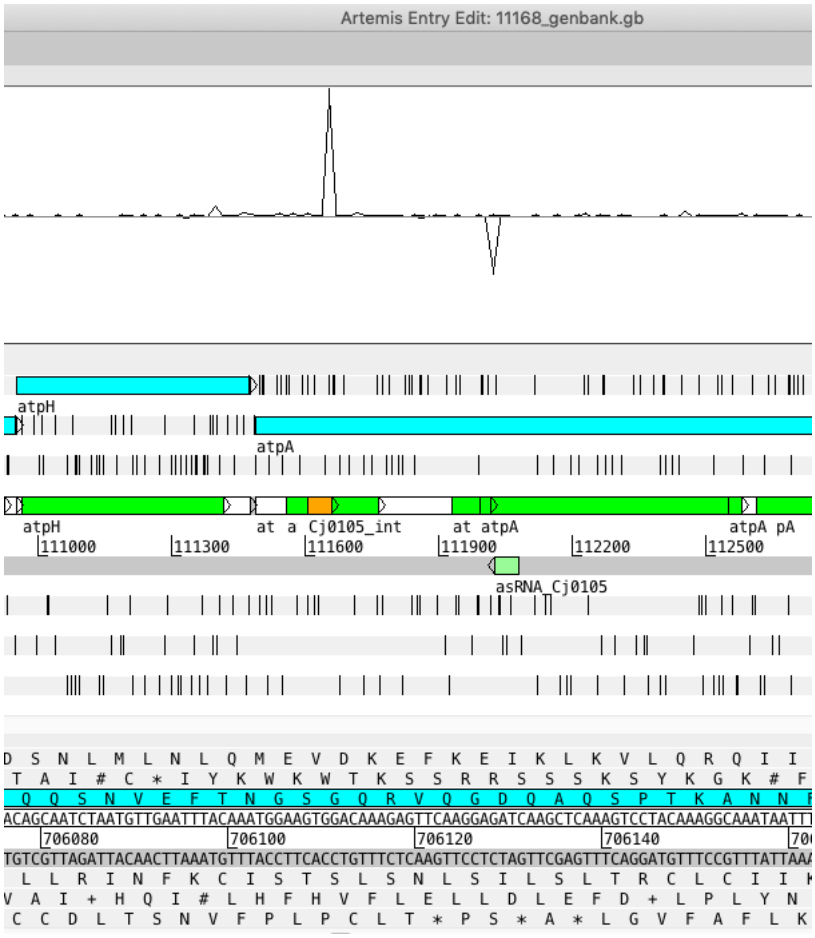


Figure 4.3: A screenshot of the coverage from the enriched raw bam file that corresponds to an internal TSS and antisense TSS on the opposite strand viewed on the Artemis genome browser. NCTC 11168 GenBank file and an annotated Artemis file with promoter information by Porcelli *et al.* (2013) was also loaded onto Artemis. The screenshot shows expression peaks for the internal annotated promoter Cj0105_int denoted by the orange box and the antisense promoter asRNA_Cj0105 of the gene *aptA*. The TSS is at the 3' end of the promoter.

For each TSS, an enrichment score was calculated representing the ratio of the RRS for the enriched sample over the RRS for the control sample. A filter was applied using the cut-off $\log_2$ fold enrichment score greater than 0.

$$\log_2(\frac{RRSenriched}{RRScontrol}) > 0$$

After filtering, the output file contains 13,253 TSS. Lastly, cluster_tss.pl was used to cluster TSS within 10 bp of each other that are likely from the same promoter and a result of "wobbly" imprecise transcription, retaining the TSS with the highest RRS (Ettwiller *et al.*, 2016). The output clustered within 10 bp gives a final total of 5,197 TSS. Figure 4.4 is a schematic flowchart visually representing the above workflow.

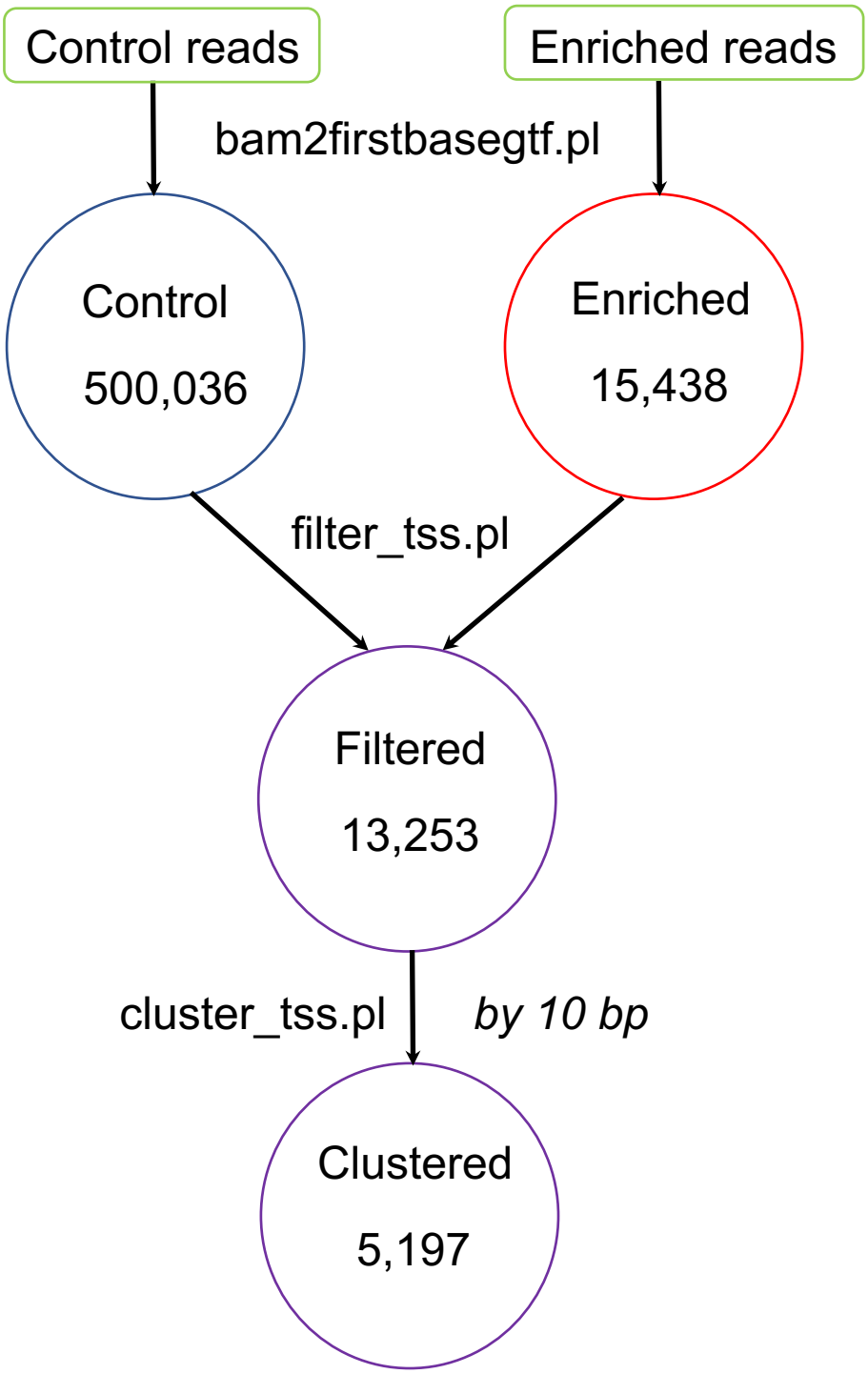


Figure 4.4: Schematic flowchart of Cappable-seq workflow. Numbers inside circles represent the number of TSS identified at each stage.

4.2.1 Comparison with published data

The two published studies on the primary transcriptome in *C. jejuni*: Dugar *et al.* (2013) and Porcelli *et al.* (2013) used dRNA-seq libraries to determine TSS and have identified 1,905 and 992 TSS respectively. Handley *et al.* (2015) also used dRNA-seq to determine TSS of four specific PerR-regulated genes and eight Fur-repressed genes giving a total of 12 identified TSS. However, on closer inspection the data set from Dugar *et al.* (2013) has multiple TSS within 10 bp of each other on the same strand which are likely from the same promoter. Thus, to make that data set comparable with our data set as described above, pairs of TSS were screened and subsequently clustered together if they were within 10 bp. As there are no RRS data available for TSS from the published data set, the earliest TSS from either strand was retained instead, reducing the original total of 1,905 in Dugar *et al.* (2013)'s data set to 1,836 positions. Both data sets from Porcelli *et al.* (2013) and Handley *et al.* (2015) were also screened using the same script and parameters but there were no TSS found that were within 10 bp of each other. Since 10 bp was also used as a cut-off to cluster TSS from our data set using Cappable-seq, a 10 bp range below and above ($\pm$10) the TSS position was used to compare all the published data sets to look for any overlap between them before comparing with our data set.

All published data sets were compared with each other with a $\pm$10 bp range to find TSS that were common in both data sets. When Porcelli *et al.* (2013)'s data set was compared against Dugar *et al.* (2013)'s study there were 792 matches. When the comparisons were switched around so that Dugar *et al.* (2013)'s data set was compared against Porcelli *et al.* (2013)'s data set, there were 793 matches. The extra match was due to the boundary set with the 10 bp range. Two TSS from Porcelli *et al.* (2013) with the positions 1294094 and 1294105 were both within the 10 bp range of Dugar *et al.* (2013)'s TSS position 1294095. Consequently, the list of 793 was chosen to make a new file containing the matching TSS, and TSS unique to each data set for comparison with Cappable-seq data. Since two TSS from Porcelli *et al.* (2013) matched to one from (Dugar *et al.*, 2013) this resulted in a total of 1,837 TSS for Dugar *et al.* (2013) within the 10 bp range. There were no overlapping TSS within 10 bp in Handley *et al.* (2015)'s data, and only one TSS that matched with Dugar *et al.* (2013), and five that matched with both Dugar *et al.* (2013), Porcelli *et al.* (2013), and our data set.

The new file combining all the published data sets was compared with Cappable-seq data to identify novel TSS. There were 897/992 matches (90.42 % overlap) with Porcelli *et al.* (2013)'s data set, 1,430/1,837 matches (77.84 % overlap) with Dugar *et al.* (2013)'s data set, and 12/14 (85.71 % overlap) with Handley *et al.* (2015)'s data set within a 10 bp range. All published data sets were combined to reveal 2,044 unique TSS. Overall there were 1,584/2,044 matches (77.49 % overlap) between Cappable-seq TSS and the published data sets. Comparing the 2,044 combined published TSS with 5,197 Cappable-seq TSS revealed 3,613 TSS, which did not match between published data and Cappable-seq data, so there are 3,613/5,197 (69.52 %) novel TSS identified from Cappable-seq. This is summarised in a Venn diagram in Figure 4.5.

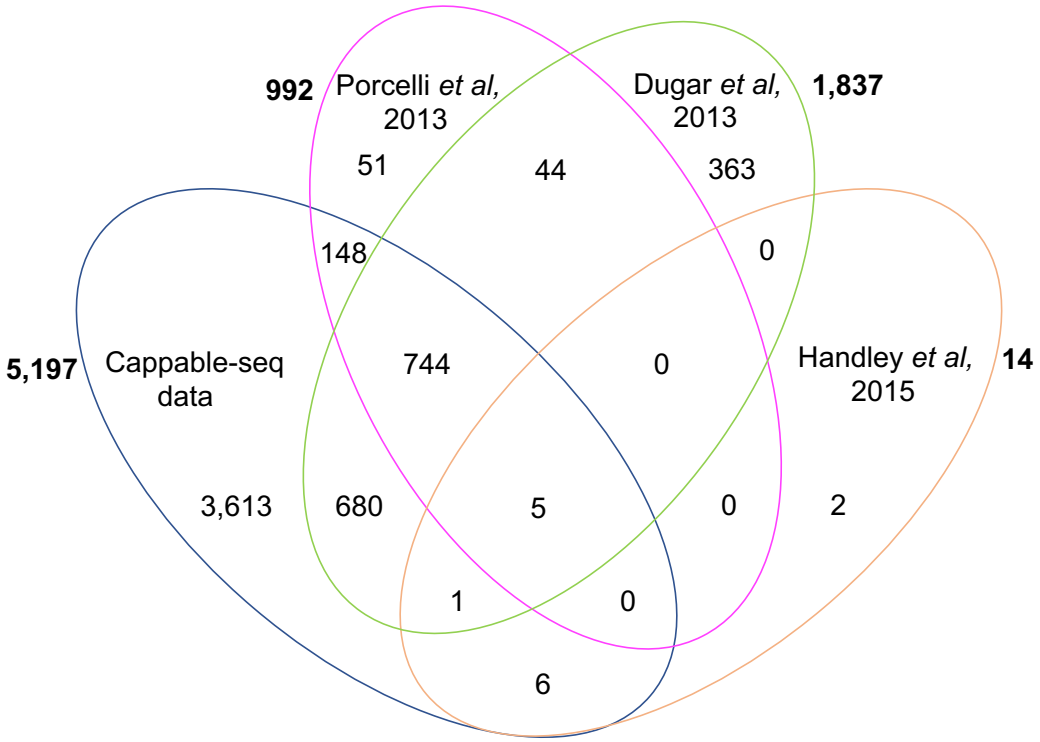


Figure 4.5: Venn diagram of TSS in each published and Cappable-seq data set. Numbers in bold depict the total TSS that were identified in each data set after filtering and clustering.

4.2.2 Significantly enriched TSS

Given the high number of novel TSS candidates obtained from Cappable-seq data, it is imperative to use a statistical test to ensure the likelihood that these are real and not artefacts. ToNER (Chapter 2, Section 2.7.3) was used as a tool to look at statistical significant enrichment between the control and enriched bam files to give high confidence to the 3,613 novel TSS found in the Cappable-seq data (Promworn *et al.*, 2017). An R-squared value of 0.8 and cut-off $p \leq 0.05$ were used as parameter thresholds. The GFF file including published ncRNAs and putative ncRNAs identified in Chapter 5 was used as an input to establish the position of TSS relative to annotated gene features. Reads were only read from the start using the option `-r start` as TSS are at the most 5' end of each read. The output from ToNER gave a very high number 55,476 of significantly enriched positions. When this was matched to the TSS found using Ettwiller *et al.* (2016)'s workflow i.e. Cappable-seq, 5,055/5,197 (97.26 %) TSS were significantly enriched. 3,479/3,613 (96.29 %) novel TSS were also significantly enriched. Eight TSS, which were found in Cappable-seq data and either Dugar *et al.* (2013) or Porcelli *et al.* (2013), were not significantly enriched with ToNER though there is a possibility that these are false-negatives and could still be genuine.

4.2.3 Promoter predictions

A promoter is the region upstream of a TSS defining where RNA polymerase binds before initiation of transcription. By predicting and locating the promoter sequence motif upstream of a TSS, it is further confirmation that the TSS is genuine and also gives information about the sigma factor preference of the gene depending on the promoter motif present. To predict promoters from the list of TSS found from Cappable-seq, the promoter region 50 nt upstream of each TSS (clustered within 10 bp) was searched for nucleotide bias to predict motif sequences. As there are three sigma factors encoded in NCTC 11168: σ^{70} (RpoD), σ^{54} (RpoN), and σ^{28} (RpoF/FliA), it was expected for three promoter motifs to be found (Parkhill *et al.*, 2000). The 5,197 sequences were submitted to MEME-suite[1] with a maximum motif size of 50 nt (Bailey and Elkan, 1994). The motif for σ^{70} and σ^{54} were identified and similar to what Dugar *et al.* (2013) and Porcelli *et al.* (2013) found (Figures 4.6 and 4.7). Although our σ^{70} motif has a strong G upstream of the A/T rich pribnow box (Figure 4.6) which was not as pronounced in the published studies. In *E. coli*, it was initially proposed there was an extended -10 element 14-15 bp upstream of the TSS consisting of the nucleotides $^{-15}$TG^{-14} but analysis of the promoters using weight matrices found a conserved G at position -13 similar to the pronounced G in our MEME search (Djordjevic, 2011). The typical A/T periodicity for *C. jejuni* was also seen instead of the usual -35 box (Petersen *et al.*, 2003). However, there was not a clear motif found for σ^{28} (Figure 4.8). The reasons for this are unknown but it is likely due to the background noise from all 5,197 sequences and/or "wobbly" transcription, as TSS within 10 bp were clustered together so not all inputted sequences had an accurate TSS and may not be aligned correctly.

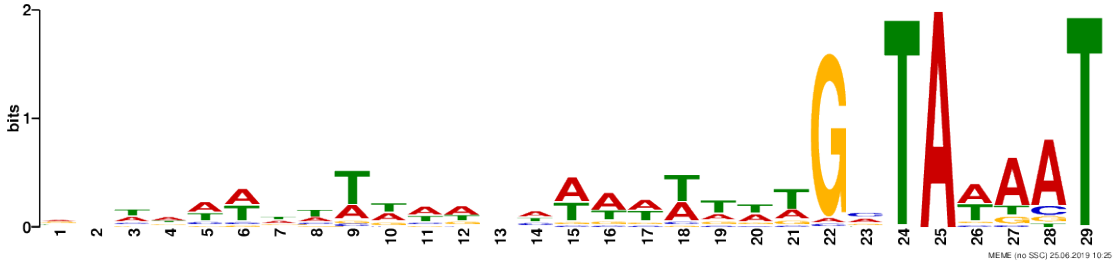


Figure 4.6: MEME logo of σ^{70} promoter motif. The size of the predicted motif is given on the x-axis. There is the typical A/T periodicity upstream of the -10 pribnow box and a pronounced G upstream.

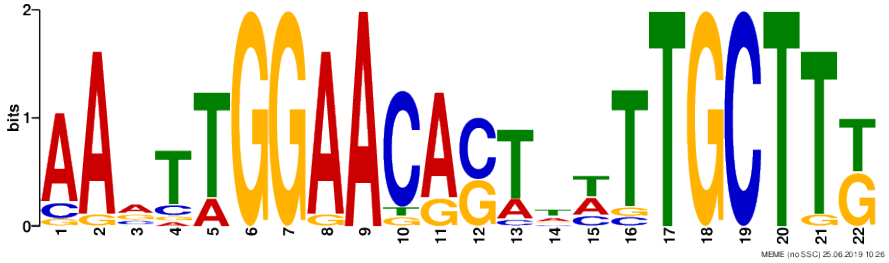


Figure 4.7: MEME logo of σ^{54} promoter motif. The size of the predicted motif is given on the x-axis.

[1] `http://meme-suite.org/tools/meme`

Figure 4.8: MEME logo of σ^{28} promoter motif. The size of the predicted motif is given on the x-axis.

The 50 nt sequence upstream of each TSS was subsequently searched for promoter motifs for each sigma factor using the R package BioStrings for string search using the matchPattern() function. As there was no σ^{28} motif available from the MEME search and the recognition sequences for the other two sigma factors were similar to what was found in Porcelli *et al.* (2013), the promoter motifs and distance restrictions from Porcelli *et al.* (2013)'s study were used to assign promoters for the Cappable-seq data as summarised in Table 4.2. Manual curation of the σ^{28} motif from our data may be necessary in the future though this was not carried out due to time constraints. The σ^{54} motif 5' GG-N9-TGCT was found from the supplementary table from Porcelli *et al.* (2013) and is a loose interpretation of the MEME logos we have found in Figure 4.7 and in the supplementary materials of Dugar *et al.* (2013). There were not as many σ^{70} promoters identified with the option of no mismatches and the set distance restriction. Therefore, sequences that had no promoter found were once again searched with the σ^{70} motif but with one mismatch allowed and these were assigned as weak promoters. It was also not clear whether a distance restriction was used in Dugar *et al.* (2013).

Sigma factor	Motif sequence	Distance upstream of TSS	Dugar	Porcelli	Handley	Cappable-seq	Mismatches
σ^{28}	5' CGATWT	6-8 nt	71	26	-	122	
σ^{54}	5' GG-N9-TCGT	8-13 nt	13	18	-	51	
σ^{70}	5' GNTANAAT	4-8 nt	1,682	948	14	905	(+1) 2,179 (+2) 1,342
		Total	1,745	992	14	1,078	4,387

Table 4.2: Table of promoter motifs identified in the promoter regions of TSS from Cappable-seq and published studies. The search for the σ^{70} motif in promoters upstream of Cappable-seq TSS was relaxed to one and two mismatches indicated by '(+1)' and '(+2)' respectively with the additional number of TSS shown.

With one mismatch, a total of 3,257 σ^{70} promoters were found upstream of the 5,197 TSS from Cappable-seq. The remaining promoter-less TSS may be from spurious transcription, although some predicted promoters identified from all three published studies also did not have any motif found using our criteria. Moreover, many novel promoters found upstream of genes without an annotated promoter in Section 4.3.2 were manually inspected and found to have a loose σ^{70} motif with two mismatches and the same was found in published promoters that also did not identify a promoter with our strict criteria. Thus, the criteria was relaxed to incorporate two mismatches giving a total of 4387/5,197 (84.41 %) promoters found upstream of Cappable-seq TSS as seen in Table 4.2.

The strength of the promoter sequence for σ^{70} motifs were categorised as: strong = no mismatch, weak = one mismatch, and very weak = two mismatches. From the list of novel and significantly enriched TSS, 28 were found to have σ^{54}, 83 σ^{28}, and 2,894 (408 strong, 1,431 weak, and 1,055 very weak) σ^{70} motif. All but one of Porcelli *et al.* (2013) and all but two of Dugar *et al.* (2013) σ^{54} promoters were also found in the Cappable-seq data. The promoters that were not found did not have an identified TSS, although strangely the two promoters from Dugar *et al.* (2013) were classified as σ^{70} in Porcelli *et al.* (2013). For σ^{28} there was less overlap, only 16/26 of Porcelli *et al.* (2013) and 34/71 Dugar *et al.* (2013) were found with the missing having not identified a TSS in Cappable-seq. For σ^{70} there was 802/948, 1,289/1,682, and 12/14 promoters found for Porcelli *et al.* (2013), Dugar *et al.* (2013), and Handley *et al.* (2015) respectively. Some promoter sequences had either σ^{70} and σ^{54}, or σ^{70} and σ^{28} motifs but these were counted as either σ^{54} or σ^{28}. TSS which have a promoter motif and were significantly enriched using ToNER were considered as high confidence.

4.2.4 Categorising TSS and the generation of an annotated Artemis file

TSS were organised into five categories using a custom made script on R: primary, secondary, internal, antisense, and orphan using parameters as stated in Figure 4.9. Initially, a TSS was considered primary if it was $\leq$ 300 nt upstream from the start of an annotated gene as was used in Dugar *et al.* (2013). As there was no distance limit specified in Dugar *et al.* (2013) for secondary TSS, these were classified if they were upstream of a primary promoter and within $\leq$ 500 nt of an annotated gene as was used in Porcelli *et al.* (2013). Internal TSS were categorised if they were within a gene, and antisense TSS if they were within a gene on the opposite strand or within 100 nt upstream of the gene feature on the opposite strand. Any TSS which did not fit the criteria for these categories were considered orphan, i.e. no association with any annotated gene features.

However, it was found that many of the TSS from Cappable-seq were in close proximity to each other upstream of a gene feature even after clustering within 10 bp. Dugar *et al.* (2013) defined the primary promoter as the closest TSS upstream of a gene. Interestingly, our data shows in some cases where TSS further away from the gene had a higher RSS/enrichment score corresponding to the number of reads mapping to that TSS compared with the closest TSS to the gene. As there can only be one primary promoter

per gene, the multiple TSS were ranked according to the enrichment score and the TSS with the highest score was categorised as primary, whilst all other surrounding TSS were categorised as secondary with the number depicting the decreasing rank of the scores in ascending order (i.e. _1 is the primary promoter with highest enrichment and so on) (Figure 4.9). In rare cases where the score was the same for two TSS in close proximity, the position was used for ranking instead where the TSS with the same enrichment score closest to the gene feature is assigned a higher rank. It is important to note that this categorisation does not take into account ncRNAs in the 5' UTR and manual examination of each promoter may be necessary to distinguish ncRNAs from transcripts with long UTRs.

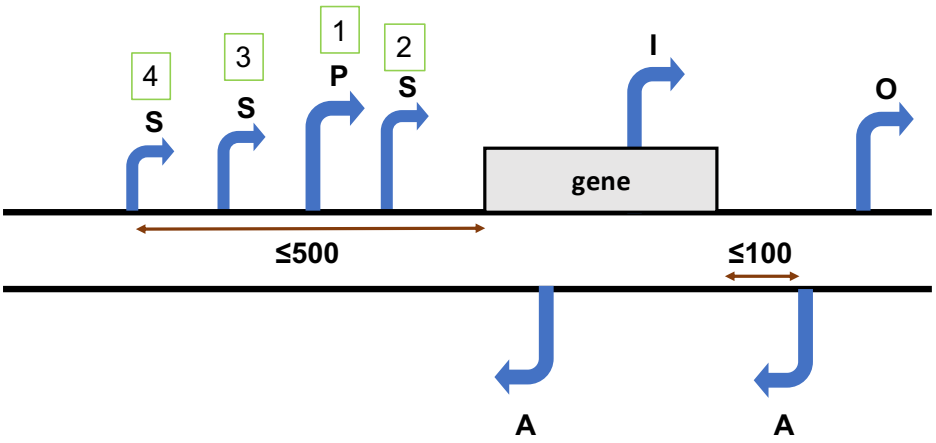


Figure 4.9: Schematic diagram for categorising types of TSS. P = Primary, S = Secondary, I = Internal, A = Antisense, O = Orphan. The primary promoter is defined as the TSS with the highest enrichment score at the 3' end of the promoter within 500 bp upstream of a gene, with other promoters defined as secondary if present. Internal promoters are defined if the TSS is intragenic, antisense promoters are defined if the TSS is on the opposite strand of a gene or within 100 bp on the opposite strand of a gene, and orphan promoters contain TSS that do not fall in the above categories. Numbers in green boxes depict a hypothetical ranking system of multiple TSS upstream of a gene feature based on enrichment score, and brown arrows depict the maximum distance from a gene feature.

78 putative primary novel TSS did not have secondary promoters, with 42 containing a promoter motif. Each TSS was given an identifier with the prefix tss_ starting from 0 from the lowest to highest genomic position. This was essential for the analysis in Section 4.3.1, but also useful for distinguishing putative ncRNA transcripts from annotated gene features.

Orphan TSS were manually inspected visually using Artemis and compared with published data. If there was an associated annotation from published data then this was updated in the file including the un-annotated selW orthologue and CRISPR components. Otherwise, the TSS were named sequentially intergenic_1 and so on in the order they appear from lowest to highest position in the genome. Of the 3,005 novel TSS which have a promoter motif and is significantly enriched, the majority were found to be internal - 1,949/3,005. The full breakdown of categorised TSS can be seen in Table 4.3.

TSS type	Total	Novel	Novel & enriched	Novel, enriched, with promoter
Primary	637	154	154	123
Secondary	317	192	176	143
Internal	2,956	2,388	2,277	1,949
Antisense	1,257	861	855	767
Orphan	30	18	17	14

Table 4.3: Table of TSS categorised into different types. Numbers were broken down into: novel (not found in published studies); novel and enriched (i.e. enriched with ToNER); and novel, enriched, with promoter (also with a promoter motif present).

All the above information from 5,197 TSS was collated into an Artemis file (found in supplementary materials) for visualisation purposes and to determine if multiple TSS upstream are novel or published. In the file, each promoter shows information on whether it was found from either published study, if it was enriched with ToNER, if any promoter motif was present, and a TSS identifier. The end position of each promoter is the TSS. The Artemis file can also show promoters that overlap with each other. It is strongly recommended to install the Artemis visualisation tool and download the Artemis file, which is available at the following GitHub repository `https://github.com/Jenna-Lam/Supplementary-material-for-thesis/blob/master/README.md` for ease of understanding the following results and discussion. The full list of TSS with their strand information, sigma factor, promoter sequence, comparison with published data, enrichment with ToNER, and categorisation can also be found in the supplementary materials.

4.2.5 Identifying ncRNAs from Cappable-seq data using toRNAdo

Since a sRNA library kit was used to generate libraries from Cappable-seq processed RNA samples we anticipate many novel putative antisense ncRNAs to be discovered from this data set. Many antisense TSS were identified from the previous section but have no defined length. Therefore, toRNAdo[2] which is a tool used for identifying transcripts in un-annotated regions was used for this purpose on the Cappable-seq data (Hermansen *et al.*, 2018). The tool searches for a 5-fold change or above in expression from a base level in un-annotated regions, therefore sequence features were removed from the GFF file prior to use as an input. This gives an output csv file containing a list of ncRNA candidates, their start position, stop position, length of ncRNA, type of ncRNAs, strand orientation, and expression peak. Only the treated bam file containing enriched 5'-PPP/5'-PP transcripts was used as an input for toRNAdo. 459 putative ncRNAs were identified by toRNAdo from the enriched bam file. Of these 371 had a TSS within 10 bp: 163 overlapped with either Dugar *et al.* (2013) and/or Porcelli *et al.* (2013), and 208 had a novel TSS. All 371 putative ncRNAs that had a TSS within 10 bp were subsequently added to the GFF file and used for normalisation with TPM and DESeq2 analysis in Chapter 5, Section 5.3.2.

[2] `https://github.com/pavsaz/toRNAdo`

4.3 Analysis of TSS

4.3.1 Conditionally expressed TSS

The same pool of RNA from 22 conditions used for Cappable-seq was sequenced individually using RNAtag-Seq in Chapter 5. This allows us to determine which conditions TSS were derived from by comparing both data sets. The protocol used was taken from Kröger *et al.* (2013) and explained in more detail in Chapter 2, Section 2.7.3. TPM was calculated for 10 nt of each TSS "transcript" per sample using bam files from RNAtag-Seq as input for coverageBed along with a custom GFF file containing all TSS "transcripts". A TSS identifier was given for each TSS starting from 0 following the format of the GFF file. TSS were counted as originating from one condition if it was present in at least two biological replicates of a condition. From the calculated TPM values of each TSS transcript, 2,170/5,197 TSS were found to be conditionally expressed meaning in at least one condition the expression was below 10 TPM in a minimum of two biological replicates. This indicates a wealth of data that can be analysed in more depth to unravel the transcriptional architecture of *C. jejuni*.

Short-hand	Meaning
E	exponential
S	early stationary
L	late stationary
a	all
b	sodium deoxycholate
IL	iron limitation
IR	iron repletion
cj100	100 % chicken juice
cj5	5 % chicken juice
37	37 °C
42	42 °C
4	4 °C
_(n)	replicate

Table 4.4: Shorthand key for conditions.

Figure 4.10 is a visual representation of all 5,197 TSS across all samples with the shorthand key for condition names in Table 4.4. Anything $\leq$ 10 TPM was treated as not expressed. 100 % chicken juice was not included as it was established in Chapter 5 that the alignment was very low with most reads mapping to the chicken genome, thus data from this condition cannot be trusted and was removed to prevent skewing the data. Some TSS had 0 TPM across all samples but it is important to consider that Cappable-seq libraries enriches for 5'-PPP/5'-PP and used sRNA library preparation contrary to RNAtag-Seq.

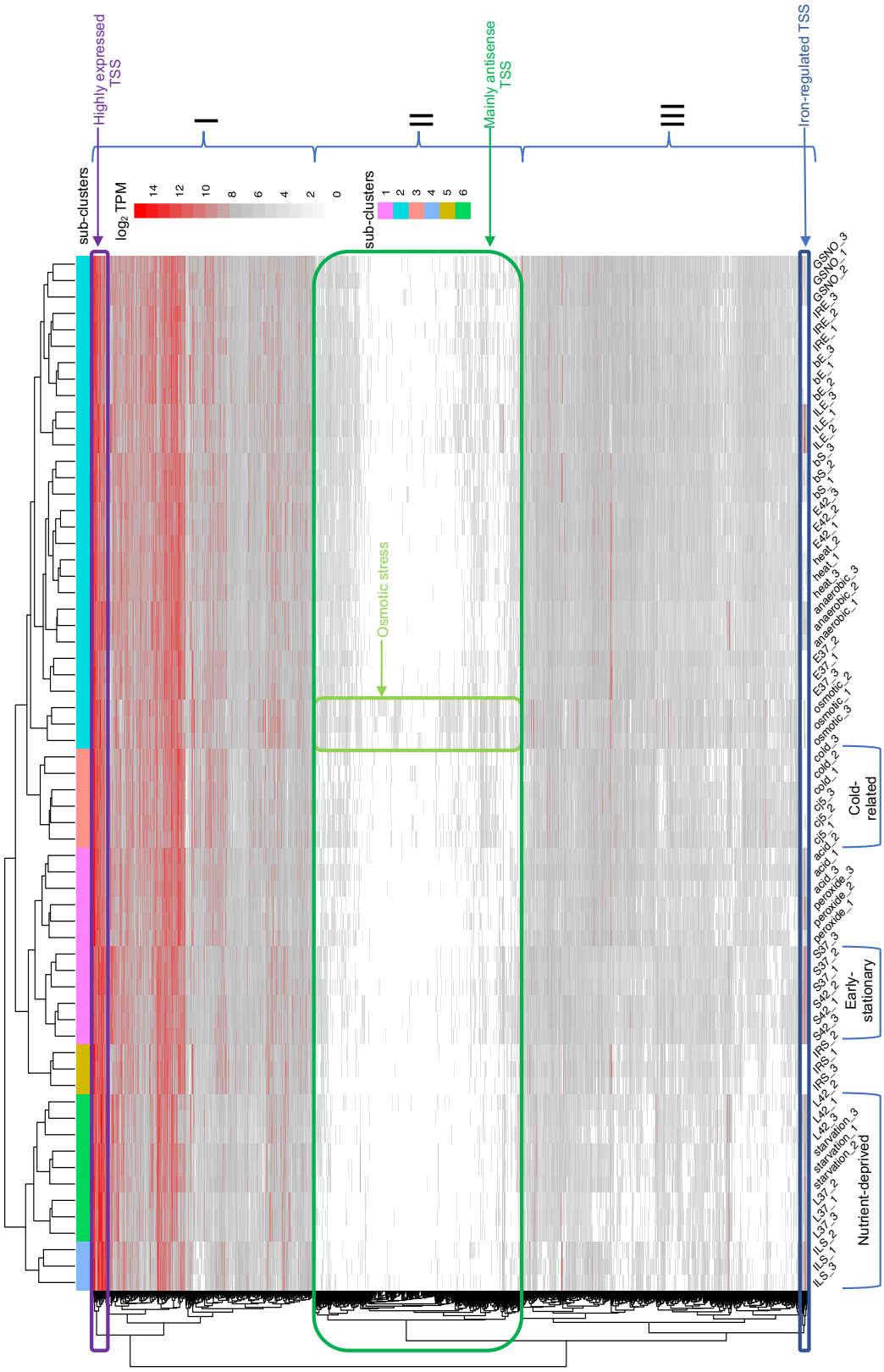


Figure 4.10: Heat map of $\log_2$ TPM values of TSS transcripts. Annotated boxes highlight highly expressed TSS in red, a cluster of mostly antisense TSS in green, and TSS associated with iron-regulated genes in blue.

All the replicates from each condition clustered together as seen at the bottom of Figure 4.10 indicating little variation of TSS "transcripts" across biological replicates. Conditions were divided into two main clusters that are further sub-divided into sub-clusters as shown by the top annotated dendrogram and TSS were divided into three sections on the right. It was interesting to see that sub-cluster 4, which includes all the iron limitation early stationary replicates (ILS) clustered with sub-cluster 6, containing the nutrient deprived conditions late stationary growth phase at both host temperatures (L42 and L37), and starvation in the left main cluster as labelled in Figure 4.10. It seems sub-cluster 4 and 6 cluster separately with all the other conditions due to the difference in expression observed in section III, where there are less TSS expressed. As ILS clusters with the late stationary and starvation conditions, it seems this stress has a similar pattern of activating/repressing genes possibly due to a similar nutrient depleted profile. The main cluster on the right contains the rest of the conditions with sub-cluster 1 and 5 together. Sub-cluster 5 contains iron repletion early stationary (IRS) replicates only, whereas sub-cluster 1 has early stationary growth phases at both host temperatures (S37 and S42) along with peroxide stress and acid shock. The remaining sub-clusters are 3 and 2, with 3 containing all the cold-related conditions and 2 all the remaining conditions.

Section 1 contains TSS which are expressed in every condition apart from a few replicates of ILS for a few TSS. Some of the TSS are also very highly expressed in each condition as labelled by the purple box. Section 2 (highlighted by the green box) contains many TSS that were only expressed in a few conditions. Most notably is osmotic stress in sub-cluster 2, which had the most expressed TSS in section 2. Many of the TSS were in fact antisense TSS possibly driving potential ncRNAs. Section 3 at the bottom has moderate expression of TSS in most conditions apart from sub-clusters 3 and 6 on the left. There is also a small cluster of TSS at the bottom labelled in blue that are highly expressed in some conditions but have no expression in others. There is expression in GSNO, iron limitation exponential (ILE), anaerobic, and sub-clusters 1, 4, 5, and 6. Many of the TSS in this cluster are internal and reside within iron acquisition genes.

4.3.2 TSS under iron-limited conditions

Iron homeostasis is an important and essential process in all living organisms including bacteria and is therefore a very well-studied aspect in microbiology. As we would like to evaluate and validate our Cappable-seq data as a potential resource and data base, iron-related conditions were looked into more deeply, focusing on genes that are iron repressed independent of growth phase. The list of TPM values for each TSS transcript that were used to generate Figure 4.10 and the list of TPM values of gene features from RNAtag-Seq were filtered for expression in the absence of iron i.e. expressed in both exponential and early stationary iron limitation conditions and absent in both exponential and early stationary iron repletion conditions (repressed by iron), using a cut-off of 10 TPM in a minimum of two replicates. 25 gene features including eight ncRNAs and one pseudogene were found to fit this criteria and were compared with TPM values derived from TSS transcripts (Table 4.5). These were also confirmed by looking at the $\log_2$ fold-change between all iron limitation vs iron repletion comparisons from DESeq2 analysis (Section 5.3.2). There were also TSS that were repressed by iron but the associated gene did not make the cut-off. These genes were expressed under iron-limited conditions but were also lowly expressed just above the 10 TPM cut-off in iron-repleted conditions. Only the genes that made the cut-off were explored in depth. It is recommended to download and install the Artemis file found at the GitHub repository[3] for ease of understanding the following results and discussion. 100 % chicken juice was not accounted for.

[3] `https://github.com/Jenna-Lam/Supplementary-material-for-thesis/blob/master/README.md`

Locus tag	Gene name	Description	TSS name	TSS type	Published/novel	Promoter	Fur-binding box
as_Cj0174c_2	-	upstream of Cj0177	tss524_as_Cj0175c*	A	published[b]	very weak σ^{70}	-
Cj0177	-	iron transport protein	tss531_Cj0177	P	published[h]	very weak σ^{70}	-
			tss532_Cj0177_int	I	published[d]	very weak[u] σ^{70}	-
Cj0178[o]	-	TonB-dependent outer membrane receptor	-	-	-	-	-
Cj0179[o]	*exbB1*	biopolymer transport protein	tss540_Cj0178_int_5	I	novel	weak σ^{70}	-
			tss539_Cj0178_int_4	I	novel	weak σ^{70}	-
			tss537_Cj0178_int_3	I	published[d]	strong σ^{70}	-
			tss535_Cj0178_int_2*	I	published[d]	very weak σ^{70}	-
			tss534_Cj0178_int	I	novel	σ^{28}	-
Cj0180[o]	*exbD1*	biopolymer transport protein	-	-	-	-	-
int_Cj0753c_1	-	upstream of Cj0177 *tonB3*	tss2232_Cj0753c_4*	P	novel	very weak σ^{70}	(repress)
Cj0755	*cfrA*	ferric enterobactin uptake receptor	tss2233_Cj0755	P	published[h]	very weak σ^{70}	repress
Cj0866;part=2/7; pseudo=true	-	pseudogene	tss2551_Cj0866p1/7_int*	I	novel	weak σ^{70}	-
			tss2552_Cj0866p1/7_int_2*	I	novel	weak σ^{70}	-
			tss2553_Cj0866p1/7_int_3*	I	published[b]	weak σ^{70}	-
as_Cj1157_2	-	antisense to *dnaX* DNA polymerase III	tss3450_as_Cj1157*	A	published[d]	weak σ^{70}	-
as_Cj1262_1	-	antisense to racS two-component sensor	tss3758_as_Cj1262	A	published[b]	very weak σ^{70}	-
			tss3759_as_Cj1262_2	A	novel	very weak σ^{70}	-
as_Cj1353_2	-	antisense to ceuC enterochelin uptake permease	tss4037_as_Cj1353*	A	published[b]	strong σ^{70}	-
			tss4038_as_Cj1353_2*	A	novel	weak σ^{70}	-

Table 4.5 continued from previous page

Locus tag	Gene name	Description	TSS name	TSS type	Published/novel	Promoter	Fur-binding box
as_Cj1385_1	-	antisense to *katA*	tss4155_as_Cj1385*	A	published[b]	very weak σ^{70}	-
Cj1384c	-	hypothetical protein	tss4150_Cj1384c_1	P	published[h]	very weak σ^{70}	-
			tss4151_Cj1384c_2	S	novel	weak σ^{70}	-
			tss4149_Cj1384c_int_2	I	novel	weak[u] σ^{70}	-
			tss4154_Cj1384c	P	novel	weak σ^{70}	-
as_Cj1383c	-	upstream of katA	tss4146_Cj1385_3*	S	published[d]	weak σ^{70}	-
Cj1385	*katA*	catalase	tss4153_Cj1385_4	S	novel	weak σ^{70}	-
			tss4152_Cj1385_1	P	published[h]	strong σ^{70}	-
			tss4148_Cj1385_2*	S	novel	very weak σ^{70}	-
			tss4146_Cj1385_3*	S	published[d]	weak σ^{70}	-
Cj1386	-	ankyrin repeat-containing protein	tss4157_Cj1386	P	novel	weak σ^{70}	-
Cj1394[o]	-	adenylosuccinate lyase	tss4172_Cj1393_int_3*	I	novel	weak σ^{70}	-
			tss4171_Cj1393_int_2*	I	novel	weak σ^{70}	-
			tss4170_Cj1393_int	I	novel	very weak σ^{70}	-
as_Cj1614_1	-	antisense to Cj1614	tss4877_Cj1614_2	A	published[d]	weak σ^{70}	-
Cj1614	*chuA*	hemin uptake system outer membrane receptor	tss4877_Cj1614_1	P	novel	weak σ^{70}	-
			tss4878_Cj1614_2	S	novel	very weak σ^{70}	repress
Cj1616[o]	*chuC*	hemin uptake ABC transporter ATP-binding protein	tss4885_Cj1615_int_3	I	novel	very weak σ^{70}	-
			tss4884_Cj1615_int_2	I	published[b]	very weak σ^{70}	-
			tss4883_Cj1615_int	I	novel	strong σ^{70}	-

Table 4.5 continued from previous page

Locus tag	Gene name	Description	TSS name	TSS type	Published/novel	Promoter	Fur-binding box
Cj1617[o]	*chuD*	hemin uptake system substrate-binding protein	-	-	-	-	-
Cj1628	*exbB2*	ExbB/TolQ family transport protein	tss4926_Cj1628	P	published[h]	weak σ^{70}	repress
			tss4927_Cj1628_int	I	novel	very weak[u] σ^{70}	-
Cj1629[o]	*exbD2*	ExbD/TolR family transport protein	-	-	-	-	-
Cj1630[o]	*tonB2*	TonB transport protein	-	-	-	-	-
as_Cj1631c_1	-	antisense	tss4934_Cj1631c_int*	I	novel	weak σ^{70}	-
			tss4931_as_Cj1631c	A	published[d]	strong[u] σ^{70}	-
			tss4932_as_Cj1631c_2*	A	novel	weak σ^{70}	-

Table 4.5: Table of genes expressed in iron limitation but repressed in iron repletion conditions both at exponential and early stationary phase. [o] the gene is found within an operon and not the starting gene. [b] found in both Dugar *et al.* (2013) and Porcelli *et al.* (2013). [d] also found in Dugar *et al.* (2013) only. [h] also found in Handley *et al.* (2015). [u] motif was found but not within the distance restriction. *denotes the TSS "transcript" was not expressed in iron limitation and repressed in iron repletion conditions at both exponential and early stationary growth phases. (repress) indicates a holo-Fur binding box was identified but within the gene feature rather than the promoter.

Many of the genes in Table 4.5 are involved with iron uptake with the exception of the pseudogene Cj0866 and Cj1386 ankyrin repeat-containing protein. The catalase gene *katA*, though not directly involved in iron uptake, is known to be regulated by the metalloregulators PerR and Fur, which are both iron dependent (Butcher *et al.*, 2015; Handley *et al.*, 2015). Cj1386 may be linked to haem trafficking for *katA*, possibly to enable proper protein folding, which might explain its close proximity to *katA* and why it is iron-regulated (Flint *et al.*, 2012). Interestingly, *katA* has four promoters found upstream of the gene: two secondary that are novel, the secondary promoter tss4146_Cj1385_3 also found in Dugar *et al.* (2013)'s study, and the primary promoter tss4152_Cj1385_1 also found in Handley *et al.* (2015). There is a possibility that all four promoters do indeed regulate *katA* but may be conditionally dependent. However, only two of the promoters tss4153_Cj1385_4 and tss4152_Cj1385_1 were iron repressed according to TPM values, and also expressed during peroxide stress with the latter agreeing with Handley *et al.* (2015)'s study. The two TSS which were not expressed: tss4148_Cj1385_2 and the published tss4146_Cj1385_3 are actually further upstream of the gene having a longer 5' UTR, although within 500 bp so was classed as secondary. However, TPM values for both TSS were found to not be expressed in all conditions. as_Cj1383c identified by Dugar *et al.* (2013) from Table 4.5 runs off tss4146_Cj1385_3 but runs into the gene *katA*. There are also two shorter transcripts that run off tss4146_Cj1385_3: as_Cj1384_1 identified by Dugar *et al.* (2013) and as_Cj1384c_2 identified from toRNAdo.

Strangely, the identified TSS for some genes were not expressed in the absence of iron, in particular antisense transcripts and internal TSS. Although it has been established that some antisense transcripts identified by Dugar *et al.* (2013) are overlapping genes and may be giving false expression, some internal TSS were also not expressed even though the genes they reside in were expressed. For example, *exbB1* has five internal promoters but tss535_Cj0178_int_2 was not in the list even though the gene Cj0178 was. But looking at the raw unfiltered TPM values shows that the promoter tss535_Cj0178_int_2 is in fact expressed during iron-limited conditions but has low expression just above the 10 TPM cut-off in iron-repleted conditions, therefore missing the boundary.

The majority of the genes (excluding ncRNAs) were also found to be repressed by iron in Butcher *et al.* (2012)'s study apart from Cj0866, Cj1386, and strangely Cj0177 and Cj0178, although the genes *exbB1* and *exbD1* downstream in the same operon were on in the study. However *tonB1*, which is further down in the operon was not on the list. *tonB1* was highly expressed under iron-limited conditions but only during exponential phase, and was also lowly expressed in iron-repleted exponential phase. Some of the genes are within an operon denoted by ^{o}, most notably the *chuACD* genes involved in hemin uptake although *chuB* was also not on the list. *chuB*, however does have high expression in iron limitation conditions and was present in the fold-change list from DESeq2 analysis, but the expression was slightly above the 10 TPM cut-off in the iron repletion exponential condition and therefore missed the boundary.

A majority of ncRNAs were also either antisense or in the vicinity upstream or downstream of genes in the list, although as_Cj1157_2, as_Cj1262_1, and as_Cj1353_2 are not associated with any of the genes. Cj1157 is the DNA polymerase III subunit DnaX, which did not have much expression according to TPM values across many conditions and only has very low expression in iron-limited conditions. However, there was no significant fold-change in the presence and absence of iron so it is not iron-regulated. Cj1262 is RacR part of the two-component system which has a role in colonising chickens. Again this gene is not iron-regulated. Finally, Cj1353 is CeuC an enterochelin uptake permease, which was expectantly up-regulated in the absence of iron. Although TPM values show it is not as highly expressed possibly due to none of the conditions containing siderophores. The fold-change between iron repletion and iron limitation at both exponential and early stationary were also looked at to confirm whether the gene features in Table 4.5 have differential expression under iron conditions. All the gene features in the list were found to have $\leq$ -1 $\log_2$ fold-change between iron-repleted against iron-limited conditions, confirming they are indeed iron-repressed apart from as_Cj1157_2, which only had significant fold-change at early stationary phase and as_Cj1353_2, which had no fold-change in either.

Only one gene in Table 4.5 Cj1386 did not previously have an annotated published primary promoter which Cappable-seq was able to identify. The *chuABCD* operon, which contains genes heavily involved in the regulation of iron uptake, does have a published annotated promoter identified by Handley *et al.* (2015) but was not found in our data. However, there were many published promoters that Cappable-seq corroborated with, proving the data set is robust. One observation was that both novel and published promoters from the list had no promoter motif using our strict criteria of one mismatch to search for the σ^{70} motif, but when the promoter region was manually searched for the motif with two mismatches many of the novel and published promoters actually had a very weak σ^{70} motif (two mismatches). It is possible that many of the novel promoters identified in Cappable-seq data with no promoter motif may have been missed due to the stringent parameters used, thus the promoter prediction for σ^{70} motifs was subsequently relaxed to two mismatches in Section 4.2.3. The remaining promoters with no motif identified were manually inspected and found to have a σ^{70} motif, but not within the distance restriction from the TSS. Two of these promoters were also published in Dugar *et al.* (2013). This is denoted by ^{u}. There was only one novel internal promoter that had a σ^{28} promoter.

The main difference between the genes, which have identified TSS from both previous and this current primary transcriptome study, was that these TSS "transcripts" also had expression during 37 °C exponential phase, whereas novel promoters did not have any TPM expression under standard growth conditions. Therefore, it is not surprising that many of the genes active in the absence of iron were found to have a novel promoter. From this we can conclude that determining the primary transcriptome under a compendium of conditions that are host-relevant or transmission related, identifies novel TSS which are conditionally expressed.

int_Cj0753c_1 is the only intergenic ncRNA on the list and runs off tss2232_Cj0753c_4 upstream of *tonB3* as seen in Figure 4.11. However, no reads mapped to the first 10 nt downstream of the TSS in any conditions. Although int_Cj0753c_1 was identified as a ncRNA from toRNAdo (Section 4.2.5), it resembles a UTR and is on top of the other identified secondary promoters for *tonB3*. But *tonB3* did not make the list for Table 4.5 as it does have TPM expression $>$ 10 for iron-repleted conditions, although the fold-change shows that its differential expression is very similar to int_Cj0753c_1. The secondary promoters that int_Cj0753c_1 overlaps were expressed in the absence of iron, so they could be contributing to reads mapping to the region that int_Cj0753c_1 encompasses. tss2232_Cj0753c_4 is also classified as a secondary promoter rather than an orphan promoter. *tonB3* is divergently transcribed from *cfrA* and all identified promoters from Cappable-seq are novel for this gene. The primary promoters of both genes, however, are opposite to each on other on their respective strands so the regulation of these genes seem to be complex. int_Cj0753c_1 was also found to be the most up-regulated gene feature in osmotic stress (Table 5.7, Chapter 5) so there is possible cross-talk between iron and other conditions.

4.3.3 Leaderless promoters

There were two instances in Table 4.5 where an internal promoter was found to overlap the beginning of the gene by 1 nt, suggesting the gene can encode for a leaderless mRNA: tss4149_Cj1384c_int and tss4927_Cj1628_int; both do not have an identified promoter motif though a weak σ^{70} and very weak σ^{70} was found respectively not within the distance restriction. Porcelli *et al.* (2013) found many leaderless mRNAs which have a 5' UTR $<$10 nt and some of the TSS even starting on the first nucleotide of the translation initiation codon, but with no recognisable aAGGa ribosome binding site (RBS). Cj1384c was not included in Porcelli *et al.* (2013)'s list of leaderless mRNAs and the promoter tss4149_Cj1384c_int contains an aAGGa site 6 nt upstream of the start site, but Cj1384c could both be a leaderless mRNA and have other canonical promoters. Since leaderless mRNAs are thought to be beneficial during stress responses, the length of 5'UTRs from Cappable-seq TSS were calculated (Porcelli *et al.*, 2013). 79 were found to have predicted UTRs 10 nt or less and 42 were actually on the first nucleotide of the initiation codon as was seen previously in Porcelli *et al.* (2013). 51/79 are novel TSS with 7 having 0 UTR length that were also enriched and have a promoter motif.

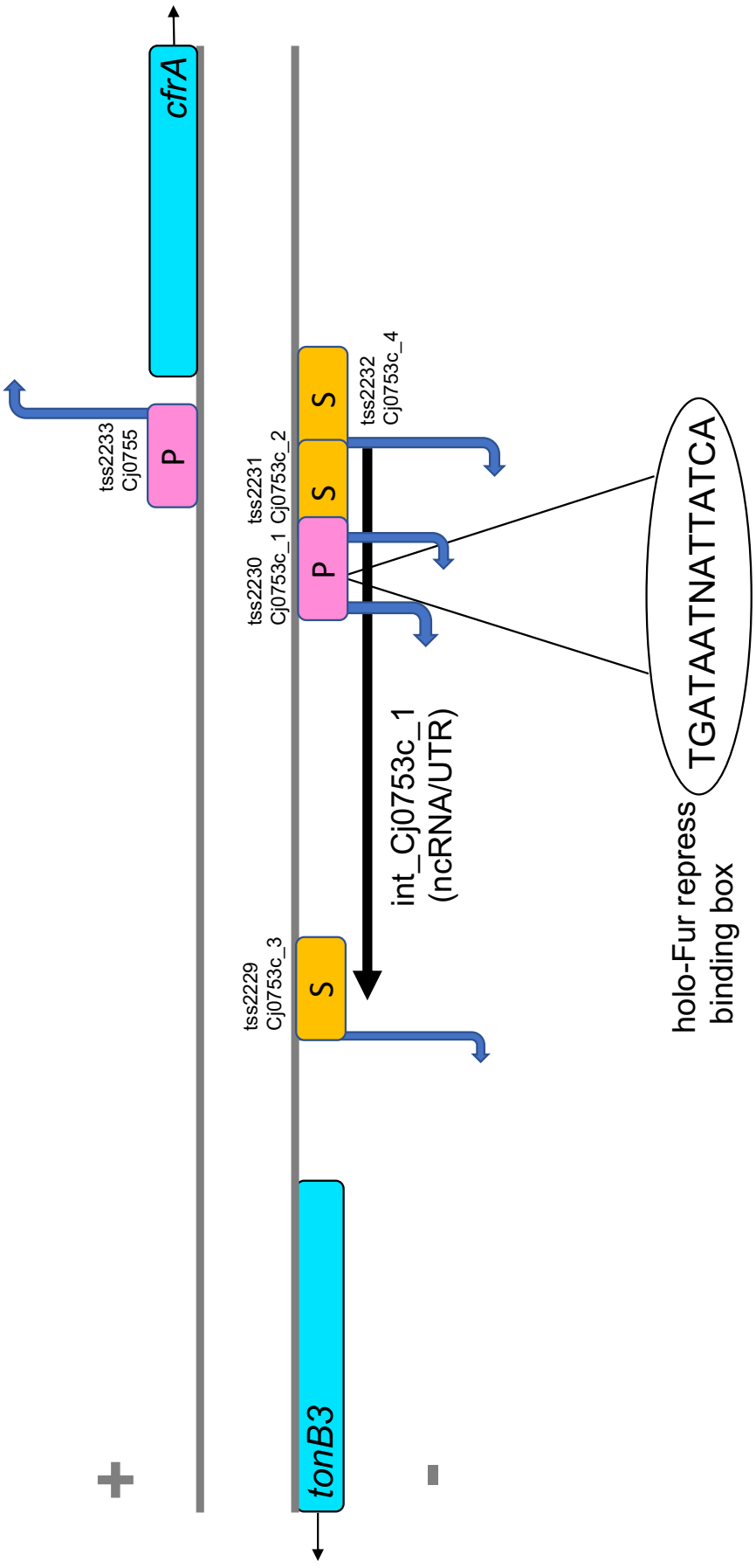


Figure 4.11: Schematic diagram of the region between *tonB3* and *cfrA*. Primary promoters are pink and secondary promoters are orange according to enrichment scores, and genes are depicted in blue.

4.3.4 Fur-binding boxes in identified promoters

Fur regulates iron acquisition genes by binding to a consensus sequence known as the Fur box and either activates or represses transcription of these genes depending on the mode of action (Palyada *et al.*, 2004). Only the holo form (Fur bound with Fe^{2+} ferrous iron) consensus sequence for activation and repression was identified from MEME in three independent studies (Palyada *et al.*, 2004; Butcher *et al.*, 2012, 2015). The Fur consensus sequences was searched throughout the genome to see if any novel promoters including the ones identified in Table 4.5, contained the consensus binding sequence using the Biostrings package matchPattern() function on R. The sequences used were as follows:

Repress: TGATAATNATTATCA (two mismatches)

Activate: TTGNNNNNNTTTNTGNT

The holo-Fur binding sequence for repression is a palindromic sequence typical for Fur boxes, but two mismatches were allowed since there are different weightings given to many of the nucleotides. The results were then compared against Cappable-seq TSS promoter regions; 16 contained repression sequences whereas 23 had activation sequences. From Table 4.5, three promoters were found to have a holo-Fur repression binding box, and one of the promoters that overlaps with the ncRNA int_Cj0753c_1 also has a Fur repression site (Figure 4.11). Tables 4.6 and 4.7 show the promoters and genes/ncRNAs they potentially regulate that have predicted holo-Fur activation and holo-Fur repression binding boxes respectively. One promoter had both holo-Fur activation and repression consensus sequence indicated by *. Internal promoters were presumed to regulate the next gene downstream, whereas antisense promoters were presumed to regulate the gene antisense to it. There were a mixture of novel and published promoters containing the Fur binding boxes.

Many of the promoters in Table 4.6 seem to regulate hypothetical proteins of unknown function such as tss179_as_Cj0044c and tss180_as_Cj0044c_2, which both regulate Cj0044c, though the two promoters overlap with each other so the holo-Fur activation sequence is actually the same in both promoters. There were two instances: tss3086_Cj1033_int_3 and tss3890_Cj1309c_int_2 where it was unclear what gene the promoter potentially regulates as both are internal promoters but there is no gene further downstream. tss3086_Cj1033_int_3 could potentially regulate an antisense ncRNA downstream of Cj1033 (annotated as *cmeF* the intergral membrane component of the CmeDEF efflux system), as there are three antisense promoters including one that was also found in published studies overlapping each other adjacent to the end of the gene. tss1678_Cj0560_int also does not have a gene downstream but the promoter placement is near the beginning of the gene Cj0560 and there is a possibility that it encodes for a short peptide or truncated version of Cj0560 as there is a RBS downstream of the promoter, although there is no clear initiator amino acid. Interestingly, the majority of promoters with a Fur box are mainly internal. tss936_Cj0321_int_3 is upstream of *perR* encoding for PerR, a global peroxide stress metalloregulator that is thought to also regulate genes in an iron-dependent manner (Butcher *et al.*, 2015). Two primary promoters for flagella genes *flaG* and *flgI* were also found to have holo-Fur activation.

TSS	Type	Published/novel	Regulates	Function
tss85_Cj0023	P	published[b]	*purB*	adenylosuccinate lyase
tss138_Cj0031_int	I	novel	Cj0033	integral membrane protein
tss179_as_Cj0044c	A	novel	Cj0044c	unknown (hypothetical protein)
tss180_as_Cj0044c_2	A	novel	Cj0044c	unknown (hypothetical protein)
tss530_Cj0175c_2	S	novel	Cj0176c	lipoprotein
tss936_Cj0321_int_3	I	novel	*perR*	peroxide stress regulator
tss1443_Cj0476_2	S	novel	*rplJ*	50S ribosomal protein L10
tss1632_Cj0547_1	P	published[b]	*flaG*	possible flagellar protein
tss1664_Cj0555_int_2	I	novel	Cj0556	amidohydrolase family protein
tss1678_Cj0560_int	I	novel	Cj0560[T]	MATE family transport protein
tss1788_Cj0604_2	S	novel	Cj0604	polyphosphate kinase
tss1822_Cj0613_4	S	novel	*pstS*	phosphate ABC transporter substrate-binding protein
tss2116_Cj0704_int*	I	novel	Cj0705	unknown (hypothetical protein)
tss3086_Cj1033_int_3	I	novel	unclear	-
tss3175_Cj1067_int	I	novel	Cj1068	zinc metalloprotease
tss3600_Cj1210_int_2	I	novel	Cj1211	competence family protein
tss3890_Cj1309c_int_2	I	novel	unclear	-
tss3928_Cj1318_int_6	I	published[b]	Cj1319	NAD-dependent 4,6-dehydratase
tss3962_Cj1333_int_4	I	published[d]	*maf3*	unknown (hypothetical protein)
tss3977_Cj1335_int_5	I	published[d]	*pseE*	unknown (hypothetical protein)
tss4024_Cj1350	P	published[p]	*mobA*	molybdopterin-guanine dinucleotide biosynthesis protein
tss4462_Cj1462	P	published[p]	*flgI*	flagellar basal body P-ring protein
tss4520_as_Cj1482c_2	A	published[d]	Cj1483c	lipoprotein

Table 4.6: List of holo-Fur activated promoters and the genes they regulate.* indicates the promoter also contains a predicted holo-Fur repression consensus sequence. [b] found in both published studies. [d] found in Dugar *et al.* (2013) only. [p] found in Porcelli *et al.* (2013) only. [T] truncated version of gene.

Most of the promoters in Table 4.7 are involved in iron regulation including iron transport uptake systems and ABC transporters. There are also a few promoters regulating membrane protein genes. A flagellar gene, *fliD* is predicted to be regulated by an antisense ncRNA as_Cj0548_2, which is novel but not regulated by an alternative sigma factor. Focusing on the genes in Table 4.5, a select few were found to contain holo-Fur binding sequences in promoters upstream of the gene. Holo-Fur repression sequences were found in published primary promoters upstream of Cj0681, *dnaE*, *cfrA*, Cj0818, Cj1099, and *exbB2* as well as upstream of the novel primary promoter of *chuA*. The TSS for *cfrA* and *exbB2* were also identified in a *fur perR* mutant in Handley *et al.* (2015) with a potential Fur/PerR operator sequence, but our scripts did not identify a holo-Fur repression sequence in the remaining 10 promoters from Handley *et al.* (2015). A promoter within int_Cj0753c_1, of tss2230_Cj0753c_1 also contained a holo-Fur repression sequence and this promoter was also expressed in iron-limited conditions, although the gene it drives *tonB3* did not make the list as it had low expression in iron-repleted conditions. Both genes were significantly down-regulated in the presence of iron at both exponential and stationary phases. Fur is able to bind directly to DNA and

activate or repress genes in its apo form (apo-Fur without iron) and holo form (holo-Fur bound to ferrous iron) (Butcher *et al.*, 2012).

TSS	Type	Published/novel	Regulates	Function
tss1643_as_Cj0548_2	A	novel	*fliD*	flagellar hook-associated protein 2
tss1772_Cj0596_int	I	novel	*fba*	fructose-bisphosphate aldolase
tss2012_Cj0681_2	P	published[b]	Cj0681	unknown (hypothetical protein)
tss2116_Cj0704_int*	I	novel	Cj0705	unknown (hypothetical protein)
tss2150_Cj0718_3	P	published[b]	*dnaE*	DNA polymerase III subunit alpha
tss2230_Cj0753c_1	S	novel	*tonB3*	TonB transport protein
tss2233_Cj0755	P	published[h]	*cfrA*	ferric enterobactin uptake receptor
tss2380_as_Cj0801	A	novel	Cj0801	probable integral membrane protein
tss2419_Cj0818	P	published[d]	Cj0818	probable lipoprotein
tss2772_as_Cj0937	A	novel	Cj0937	probable integral membrane protein
tss3013_Cj1015c_int	I	novel	*livF*	branched-chain amino acid ABC transporter ATP-binding protein
tss3218_Cj1085c_int_3	I	novel	Cj1084c	possible ATP/GTP-binding protein
tss3259_Cj1098_int	I	published[d]	Cj1099	probable peptidase
tss4422_Cj1447c_int_3	I	novel	*kpsE*	capsule polysaccharide ABC transporter permease
tss4878_Cj1614_2	P	novel	*chuA*	hemin uptake system outer membrane receptor
tss4926_Cj1628	P	published[h]	*exbB2*	ExbB/TolQ family transport protein

Table 4.7: List of holo-Fur repressed promoters and the genes they regulate. * indicates the promoter also contains a predicted holo-Fur activation consensus sequence. [b] found in both Dugar *et al.* (2013) and Porcelli *et al.* (2013). [d] found in Dugar *et al.* (2013) only. [h] found in Handley *et al.* (2015) only.

A study by Butcher *et al.* (2012) using CHiP-chip (Chromatin immunoprecipitation with DNA microarray chip) looked at protein-DNA interactions using Fur under iron-limited and iron-repleted conditions and compared the targets with Palyada *et al.* (2004)'s study, which looked at the transriptome profile of a Fur mutant using microarrays, and Holmes *et al.* (2005), which also included proteomic analysis. The four promoters containing a Fur binding repression sequence are all novel and regulate genes, which were also found to be repressed by iron in published studies (Butcher *et al.*, 2012). However, there were some conflicting results. tss_4877_Cj1614_1, which is a novel primary promoter upstream of *chuA*, was the only promoter which agreed with Butcher *et al.* (2012)'s study that also found this gene to be a direct target of CHiP-chip. On the other hand *exbB2* was not shown to be a CHiP-chip target of Fur and was listed as having holo-Fur activation, but our promoter has a predicted holo-Fur repression site. tss2233_Cj0755 upstream of *cfrA* and tss2230_Cj0753c_4 within int_Cj0753c_1 were also repressed by holo-Fur but were not direct targets of Fur by CHiP-chip.

4.4 Overview of trends and patterns observed in promoters

4.4.1 Multiple promoters

Two genes: *folD* a bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase, and Cj0965c an acyl-CoA thioester hydrolyase had the highest number of multiple TSS (six), although some of the promoters have previously been identified for both genes. There was previously a sole promoter tss2514_Cj0855_4 identified for *folD* by both published studies. Looking at the TPM values for the TSS "transcripts", tss2514_Cj0855_4 had expression in many conditions but some of these conditions did not make the cut-off $>$ 10 TPM in two replicates i.e. no expression. tss2514_Cj0855_4 was not expressed in ILS, cold stress, IRS, L37, L42, peroxide stress, and starvation. The remaining five promoters are novel, including tss2511_Cj0855_1 designated as the primary promoter as it has the highest enrichment score. One of the promoters tss2512_Cj0855_3 did not have an identified promoter motif, whereas the rest had a mixture of very weak to strong σ^{70} motifs and slightly overlapped with each other. Interestingly, tss2511_Cj0855_1, tss2512_Cj0855_3, and tss2510_Cj0855_5 only had low TPM expression in osmotic stress in two replicates but had no expression in all other conditions, suggesting they may be alternative promoters with conditional expression. The other two secondary promoters did not have any expression. TPM values and fold-change for genes were calculated from mapped reads in Chapter 5. TPM values for gene *folD* had expression across all conditions with expression in osmotic stress in the 9th decile after partitioning the values by 10. However, there was no significant fold-change between osmotic stress and E37 for *folD*. MTHFR, the human homologue of *folD*, was also found to have multiple TSS upstream, so it is probable these TSS are genuine but was not expressed very well in RNAtag-Seq (Tran *et al.*, 2002).

Two of the identified promoters for Cj0965c were previously identified by both published studies but were ranked second and fourth. The primary promoter tss2855_Cj0965c_1 overlaps with the published promoter tss2856_Cj0965c_2. Both had TPM expression across many conditions apart from L42 and cj100. However, there were two conditions where tss2855_Cj0965c_1 had expression but not tss2856_Cj0965c_2 - starvation and bE (sodium deoxycholate exponential). The furthest secondary promoter from the gene tss2860_Cj0965c_3 had no expression across all the conditions whereas the remaining three had expression but in a few conditions. Cj0965c was consistently expressed across all the conditions (apart from 100 % chicken juice).

4.4.2 σ^{54} and σ^{28} promoters

The majority of promoters in *C. jejuni* are σ^{70} regulated, but novel and significantly enriched TSS were identified from the Cappable-seq data for the alternative promoters σ^{54} and σ^{28}. There were 28 identified σ^{54} promoters that were novel and enriched, but of these there was only one primary, one secondary, and one antisense promoter. The remaining promoters were all internal and none were orphan. However, many of the internal promoters also had σ^{70} motifs. Only three internal promoters had a σ^{54} motif exclusively. There was also one primary, one secondary, and one orphan σ^{28} promoter out of 83, although the primary and orphan promoter also had a weak and strong σ^{70} motif respectively. There were 22 antisense σ^{28} promoters with the remaining promoters all internal. It is important to note even with a 10 bp range, primary and/or secondary promoters may be classed as novel, but in actual fact there may be a published promoter in the vicinity upstream of a gene that is not within the 10 bp range. That does not mean to say that the identified primary and secondary promoters are not novel, but is not the sole promoter for that particular gene feature. The associated genes from the list of novel σ^{54} promoters were submitted to the STRING database[4] and were enriched for the Gene Ontology (GO) components: periplasmic space, cell envelope, and bacterial-type flagellum basal body, distal rod. In addition, Molybdopterin oxidoreductase Fe4S4 domain from the SMART (Simple Modular Architecture Research Tool) protein domain database was enriched. The genes were also submitted to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and were found to be part of many Metabolic Pathways. No significant enrichment was found for associated genes from the list of novel σ^{28} promoters, but many genes were part of Metabolic Pathways in the KEGG database rather than being flagella-related apart from *flhA*, *motB*, and *rpoN* (encoding for σ^{54}).

The only identified novel primary σ^{54} promoter is tss171_Cj0042 which drives *flgD*, a possible flagellar hook assembly protein. tss171_Cj0042 overlaps with another novel promoter tss172_Cj0042_int classed as internal, although it could possibly be a leaderless promoter for *flgD* as the TSS is on the first start codon but has no clear promoter motif. Both promoters are expressed across all the conditions with tss172_Cj0042_int having higher expression due to the position of the TSS, as the 10 nt downstream runs into the gene. The only novel secondary σ^{54} promoter also has a very weak σ^{70} motif (tss2042_Cj0687c_3) for the gene *flgH*, a probable flagellar L-ring protein precursor. However, *flgH* has four putative promoters upstream and tss2042_Cj0687c_3 actually overlaps with tss2041_Cj0687c_4, which has a very weak σ^{70} motif. The furthest secondary promoter is tss2045_Cj0687c_2, which was also found in both primary transcriptome published studies and has a very strong σ^{70} motif. It overlaps with the primary novel promoter tss2044_Cj0687c_1, which has a very weak σ^{70} motif. Both tss2042_Cj0687c_3 and tss2041_Cj0687c_4 are expressed across all conditions, although both are closer to the gene. Whereas the published promoter tss2045_Cj0687c_2 and the primary promoter tss2044_Cj0687c_1 did not have any expression across the conditions suggesting the other two promoters are more genuine. There was also one novel antisense σ^{54} promoter tss4802_as_Cj1585c that also has a strong σ^{70} motif antisense to Cj1585c a probable oxidoreductase. However, there was no ncRNA identified from toRNAdo for

[4] `https://string-db.org/cgi/input.pl?sessionId=LPYUoO1eV1aq&input_page_show_search=on`

this antisense promoter and TPM values show there was no expression across all conditions. The three σ^{54} internal promoters reside in the genes encoding for Cj0434 a phosphoglycerate mutase, Cft ferritin, and Cj0846 an integral membrane protein. In fact the internal promoter for Cft partially overlaps the beginning of the gene by 16 nt but also contains a RBS so it is unclear whether it is a leaderless promoter.

The novel σ^{28} primary promoter, which also has a very weak σ^{70} motif and a published secondary promoter downstream drives the gene Cj1075, a previously unknown gene that has now been re-annotated as FliW, the flagellar assembly factor. Due to the hierarchical nature of flagellar assembly, it is surprising that σ^{28}, which normally mediates the late flagellar genes, possibly regulates the early *fliW* gene. However, according to TPM values, this promoter is not expressed in any condition. There was also only one novel σ^{28} secondary promoter tss4164_Cj1387c_2 for gene Cj1387c, a previously unknown hypothetical protein now characterised as a PAS helix-turn-helix domain that regulates flagella-flagella interactions, with an annotated σ^{70} published primary promoter tss4162_Cj1387c_1 further downstream. tss4164_Cj1387c_2 has no expression across all conditions according to TPM values but surprisingly neither did tss4162_Cj1387c_1, apart from in anaerobic shock where it was very lowly expressed. The gene Cj1387c was expressed in many conditions but also had fairly low TPM values. There was also one novel orphan promoter tss833_intergenic_4, which had a σ^{28} motif but also a strong σ^{70} motif. tss833_intergenic_4 is antisense to the published promoter tss834_Cj0279_1, which drives CarB a probable carbamoyl-phosphate synthase large chain. tss834_Cj0279_1 was expressed in all conditions whereas tss833_intergenic_4 is not expressed in all conditions according to TPM values. There is a possibility of promoter clashing between tss834_Cj0279_1 and tss833_intergenic_4 due to the arrangement of promoters.

4.4.3 Promoters with two sigma factors

Some of the TSS were assigned two promoter motifs with both instances including the σ^{70} motif and either σ^{54} or σ^{28}. These were mainly internal TSS with only a few that were antisense or secondary. Some annotated published promoters were also found to have two promoter motifs. It would be interesting to investigate these further to see whether they are under the control of both sigma factors and are conditionally expressed. Indeed, some flagellar genes are dependent on the control of two sigma factors but from two different promoters in an operon, for example *flgM* in Wösten *et al.* (2010a) which shows a primary σ^{28} promoter and primary σ^{54} promoter for the operon upstream before *flgI*. Both these promoters were identified in published studies and Cappable-seq, although there were a few internal σ^{70} promoters which are novel. This is also the case for the hypothetical protein Cj1387c where Cappable-seq has found a novel secondary σ^{28} promoter tss4146_Cj1387c_2 further upstream of an already annotated σ^{70} promoter tss4162_Cj1387c_1, which was classed as primary as it has a higher enrichment score. Although the secondary promoter was picked up by Cappable-seq, there was no expression in all conditions when calculating for TPM values using read counts derived from RNAtag-Seq data. A ncRNA as_Cj1388 was identified by toRNAdo running off the secondary promoter tss4146_Cj1387c_2, but this also had no expression from TPM values. As RNAtag-Seq libraries do not enrich for primary transcripts unlike

Cappable-seq libraries it is possible that this promoter and corresponding ncRNA was not detected in RNAtag-Seq, but further experiments are needed to validate their existence. Cj1387c was shown to be important and involved in autoagglutination of flagella with a role in repressing the divergently transcribed Cj1388, which as_Cj1388 is antisense to (Reuter *et al.*, 2015). In the *C. jejuni* strain 81-176-DRH212, the homologue of Cj1387c named as HeuR was also found to regulate genes involved with oxidative stress and iron-regulation (Johnson *et al.*, 2016).

4.5 Discussion

4.5.1 Identification and categorisation of novel TSS compared with published data

Cappable-seq was chosen as the method to determine the primary transcriptome of *C. jejuni* NCTC 11168 under a pooled set of 22 host-relevant and transmission conditions. This method uses direct enrichment of 5'-PPP/5'-PP primary transcript ends rather than relying on degradation of processed transcripts, which carry the risk of incomplete degradation, and also claims to have single base resolution (Ettwiller *et al.*, 2016). Cappable-seq was successful in enriching for primary transcripts, adding to the growing repertoire of TSS and promoter regions, and revealing novel features in the transcriptional landscape of *C. jejuni*. Single base resolution was achieved but the phenomenon of imprecise "wobbly" transcription where promoters initiate transcription from closely spaced TSS within a promoter during low levels of transcription was demonstrated, as was seen in previous studies (Ettwiller *et al.*, 2016; Haberle *et al.*, 2015). Precise resolution of TSS is important as the distance between the -10 promoter motif and the TSS is very sensitive. The raw data obtained from Cappable-seq is available but for ease of comparison with published data, identified TSS were clustered within 10 bp of each other with the highest RRS retained.

When compared with the two published transcriptome studies by Dugar *et al.* (2013) and Porcelli *et al.* (2013), and a specific study that identified novel TSS by Handley *et al.* (2015) all carried out in *C. jejuni* strain NCTC 11168 to date, there was very good overlap with Cappable-seq data (Figure 4.5). Apart from the difference in method, Porcelli *et al.* (2013) and Handley *et al.* (2015) also used Roche 454 pyrosequencing, a different sequencing platform. There are advantages and disadvantages of using one system over another and depending on the time of publication, some systems may not have been available. Roche 454 pyrosequencing has disadvantages of errors in sequencing homopolymeric tracts, are prone to indels, and has lower depth of sequencing compared to the Illumina platform, although it tends to have longer read lengths (Luo *et al.*, 2012; Yang *et al.*, 2013; Levy and Myers, 2016). However, Roche 454 pyrosequencing platform has been discontinued (Levy and Myers, 2016). Apart from the importance of achieving single base resolution, *C. jejuni* has many repeat regions in its genome which exacerbates the problem (Parkhill *et al.*, 2000). The Illumina systems do have a disadvantage of variable quality from tile to tile, although this can be overcome by filtering for high quality reads post-sequencing. It was also observed that some TSS were off by 1-2 nt relative to the published positions, which was why the 10 bp range was implemented to make the studies comparable, especially as Cappable-seq data was also clustered.

Although it was expected that Cappable-seq would find many additional TSS than studies carried out using dRNA-seq, given the increased sensitivity and the fact that the samples were from a pooled repertoire of 22 different growth and stress conditions, it is astonishing that 69.52 % (3,613) of the total TSS (5,197) identified are novel. Because there was a high number of novel TSS found, a statistical test to assign confidence is needed. ToNER is a tool which was specifically designed to analyse Cappable-seq data (Promworn *et al.*, 2017). 97.26 % of the novel TSS (5,055) were found to be significantly enriched from

ToNER indicating that the high number is more likely due to the increased sensitivity of Cappable-seq rather than false-positives. There were also a few TSS found in published and Cappable-seq data, which were not significantly enriched using ToNER, so it is probable that there were some false-negatives using this tool, but the numbers were still very low. However, 66.09 % (2,388/3,613) of the novel TSS were classed as internal, so it is possible that a number of TSS identified were actually due to spurious transcription (Wade and Grainger, 2014). There was also a high number of antisense ncRNAs identified which may have been attributed to the RNA extraction method used.

Searching for promoter motifs upstream of these novel TSS would strongly support whether they are genuine or pervasive, as a promoter sequence is crucial for RNA polymerase binding. Promoter-like sequences could arise from point mutations so there is a possibility that transcripts from identified TSS with a promoter motif can potentially be spurious, but are offset by rapid degradation or premature termination (Lybecker *et al.*, 2014). From our data, two promoter motifs were found for the sigma factors σ^{70} and σ^{54}. Strangely, Dugar *et al.* (2013) and Porcelli *et al.* (2013) were both able to find a σ^{28} motif but no clear motif was found in our data. There are not many genes with a σ^{28} promoter in *C. jejuni*, as this promoter is mainly involved in expression of late flagella genes, so it is possible that the large amount of sequences inputted into MEME had too much noise. In Chapter 5, the Flagella assembly pathway was significantly activated during early and late stationary phase, but the majority of stress and growth conditions were carried out at 37 °C exponential phase and as there were roughly equal amounts of RNA from each sample pooled together, it is possible the proportion of transcripts induced by late stationary phase and involved with σ^{28} regulation was a lot lower in comparison with all other conditions at the other growth phases. This could explain why it was difficult to find a σ^{28} motif in our data set.

In addition, the concept of "wobbly" transcription may have affected the inputted sequences. 50 bp sequences upstream from the identified TSS from Cappable-seq was used as input for MEME, but it is important to note that when clustering, the TSS with the highest RSS within 10 bp was retained. This was done for ease of analysing our data but this does not mean TSS with lower RSS are not genuine, so some of the sequences may have been misaligned affecting the MEME tool to search for nucleotide bias. To account for this, manual curation of the sequences may be necessary to find the σ^{28} motif, however, due to time constraints this was not carried out but may be a future prospect. Nevertheless, σ^{28} promoters were found in our data set using the σ^{28} promoter motif from Porcelli *et al.* (2013)'s study, both novel and overlapping with published promoters, demonstrating our data does indeed identify σ^{28} associated TSS but with an emphasis on the need to correctly align sequences to identify the σ^{28} motif.

Initially, there were many promoter regions upstream of the identified TSS that did not find any promoter motif(s) present with the option of no mismatches for the σ^{70} motif (Table 4.2). Although this could be due to the strict distance range from the TSS, especially given that the list of TSS were already clustered within 10 bp, so many could have been missed due to the boundaries set. Even so, there were fewer σ^{70} motifs found compared with published data even though Cappable-seq found many more TSS and it is

expected that the majority are σ^{70} regulated, as it is a housekeeping sigma factor. Considering there were published TSS that were also found in Cappable-seq but not assigned a promoter motif by our scripts, yet all published TSS have a promoter motif annotated, the σ^{70} motif was relaxed to one mismatch. The number of TSS with a promoter found increased drastically, although there were still ≈37 % remaining that were promoterless including published promoters (Table 4.2). Surprisingly Cappable-seq found many more σ^{28} and σ^{54} motifs that are mainly involved with flagellar biogenesis than the published primary transcriptome data sets although there was not much overlap with Dugar *et al.* (2013). Most of the novel TSS regulated by these alternate sigma factors were mainly internal and only one novel primary and one novel secondary promoter for each σ^{28} and σ^{54} was found in our data.

There were some minor noticeable errors in the analysis, although due to time constraints these were not rectified as they do not change the interpretation of results. A customised script was used to categorise TSS based on the rules detailed in Section 4.2.4. The way the GFF file is constructed and how the script runs meant that some genes in an operon will both fulfil the requirements for categorisation, therefore the naming of promoters may be inconsistent and/or incorrect. The script was written to ensure existing annotation is not overwritten but in the case of the antisense (negative) strand, since the position of genes in the GFF file are in the forward direction regardless of strand, this is problematic. Nevertheless, it is clear to see which genes the promoter is regulating by looking at its placement visually on Artemis.

Many sole primary promoters for genes were manually inspected and found to have a loose TANAAT box, for example, *cfrA* (see Figure 4.2, Section 4.3.2). There was also no weighting given to the nucleotides in the motif so it is very likely that many promoter motifs were missed, which is why the search for σ^{70} motif was further relaxed to two mismatches and all TSS were included in the Artemis file regardless of the identified motif, so that potential promoters can be investigated more thoroughly. This gave a total of 4387 TSS that have a promoter motif, so 84.41 % of 5,197 TSS (Table 4.2). Hence, TSS that corresponds with ToNER and has a promoter motif can be considered as high confidence. There were some genes from Table 4.5 that did not have a promoter motif found including published ones, but on closer inspection there was actually a motif present but not within the distance criteria. However, these will have to be judged on a case by case basis, since the further the motif is from the TSS the more likely RNA polymerase is unable to bind and thus may not be genuine.

It is very plausible that the high number of TSS were actually spurious transcripts, which may have been triggered from some of the growth conditions used given the vast number that were internal. The notion of spurious or pervasive transcripts are non-canonical transcripts which are not annotated genes and can either be antisense or internal (Wade and Grainger, 2014). But as there is no clear defined function given to pervasive transcripts, it is implied that these transcripts could just be "biological noise" or "dark matter" (Wade and Grainger, 2014). In both published and our data sets, 637 TSS did not find a promoter motif and 380 of these were internal. However, these TSS could still be genuine and further validation is needed to confirm their authenticity. *C. jejuni* is especially compact with many genes in an operon; it may be possible

that the majority of the internal TSS are another level of control for expression of specific genes within an operon rather than expressing the entire operon, especially under different conditions. One phenomenon to consider is the possibility of overlapping genes reviewed by Shcherbakov and Garber (2000). There are three groups of overlapping genes consisting of: N- C- terminal fragments overlapping each other, "out-of-phase" gene overlapping where the genes are in different frames, and "in-phase" gene overlapping where the genes are within the same frame (Shcherbakov and Garber, 2000). It is highly possible that *C. jejuni* exhibits this phenomenon and as a result has many internal promoters. With this data, it is possible to explore and see if genes within an operon are expressed differently to each other using TPM values under different stresses. However, *E. coli* possesses a histone-like nucleoid structuring (H-NS) protein not found in *C. jejuni* that represses intragenic promoters, so the high number of internal promoters could also be due to the absence of H-NS (Singh *et al.*, 2014).

There were only 154 novel TSS that were classed as primary and after filtering for promoter motifs and enrichment with ToNER, there were actually only 123 high confident primary novel TSS. This number seems more reasonable as the NCTC 11168 genome is fairly small and compact, so it was not expected to have very many additional promoters. Furthermore, only 55 novel high confident TSS were the sole primary promoter for genes. After calculating absolute expression for each condition and deducing from whence TSS were expressed from, it was observed that a majority were actually expressed under standard growth conditions (Figure 4.10). This could explain the low number of primary novel TSS that were identified. Some may not be novel for a gene feature as the TSS may not be within 10 bp of a published TSS further up or downstream. From such a small and compact genome, 55 novel primary promoters are not unreasonable. Since many TSS were also expressed under standard conditions and across all conditions this suggests that only a minority of TSS and promoters are expressed under certain conditions. Even so, 2,170 TSS were found to be conditionally expressed across all conditions bar 100 % chicken juice. It would be interesting to see if there is any commonality between TSS that are only expressed under certain conditions.

It was interesting to note that some genes had multiple TSS with a promoter motif upstream of the annotation. These clusters of closely spaced multiple promoters for a gene (which can be termed promoter islands) can have a combination of house-keeping and alternative sigma factor preferences, or a single promoter that has both σ^{70} and either σ^{28} or σ^{54} motif, which has been observed previously in *E. coli* (Panyukov and Ozoline, 2013). In *E. coli*, findings suggest that RpoS (also known as σ^{38}) negatively regulates σ^{70} by competing for shared promoters (Cho *et al.*, 2014). Future work could perhaps look for patterns between the alternative sigma factors in *C. jejuni* to see whether the regulated genes are highly or lowly expressed, and whether they are condition-dependent. Some promoters did not have TPM expression in any condition, so there is a possibility that they could be false negatives. But they can also perhaps hold onto RNA polymerase once it is bound to the promoter sequence but in a "closed" complex, or be a "cryptic" promoter, which are normally inactive but can be activated by a point mutation under different stresses (Panyukov and Ozoline, 2013). This is because there are a number of finite RNA polymerase molecules

and multiple promoters may be a way to compete for RNA polymerase to ensure expression of the gene. One caveat to consider is that promoters placed nearer the gene would inevitably have more coverage, especially if the secondary promoters upstream are also expressed. However, TPM may not be the best method to determine the expression of promoters, as they could be lowly expressed but have subtle changes depending on the condition, which may still lead to big changes. Since 10 nt downstream of each TSS was used for coverageBed, perhaps the read counts could be analysed by DESeq2 to also look for fold-changes between TSS "transcripts". In addition, these multiple promoters or promoter islands could also potentially transcribe very short RNA transcripts that are terminated prematurely (Panyukov and Ozoline, 2013). Whether these RNA transcripts are ncRNAs with a potential function or very short transcripts that run into the gene remains to be seen. *C. jejuni* lacks many of the known general regulators in other bacterial species, therefore it may rely on multiple promoters and internal promoters for some form of regulation.

For this project, enrichment score was used to categorise primary and secondary promoters as a simple approach, so in instances where there are multiple promoters upstream of a gene, the highest enrichment score relative to the highest number of mapped reads for that TSS was classed as primary. This does not take into account that the other promoters classed as secondary may have a stronger (i.e. more closely matched) σ^{70} motif and/or were also found in published studies. Context is important since multiple promoters may act differently under different conditions as demonstrated by the promoters upstream of *katA*, so this should be taken into consideration when Cappable-seq data is used as a resource or for further analysis. Unfortunately, it is only possible to determine the RRS of each TSS under each individual condition if the RNA samples had not been pooled and instead were made into individual libraries and sequenced separately. Therefore, even though the TSS of a promoter may have a very high enrichment score relative to other secondary promoter(s), it is possible that only under a particular condition it is highly expressed but not necessarily in all or most conditions. If the definition of primary promoter is the one that is most commonly or constitutively used, RRS and hence enrichment score may not be the best option for differentiating this. Another concern is the definition of strong or weak promoters. Here strong indicates a perfect match to the consensus sequence for the σ^{70} motif and weak for one mismatch and so on, and a perfect match would imply that RNA polymerase binding is more efficient for that promoter. But a consensus sequence is only found as the most common conserved sequence from a set of sequences, but that may not necessarily mean the promoter will be expressed the most strongly. Highly expressed or changing genes could be explored further in the future to see whether the consensus sequences holds true for standard conditions only, stress conditions, and/or across all conditions.

4.5.2 The architecture of NCTC 11168 under iron-repressed conditions

Such a large data set is overwhelming to interpret so by focusing on a particular aspect of *C. jejuni* microbiology, it is possible to decipher the intricacies and evaluate the robustness of the data and see whether it corroborates with published studies as well as discovering new components. Iron homeostasis was chosen for further analysis as this is a well established process in many bacteria including *C. jejuni*, with available transcriptomic RNA-seq and microarray data. Yet the entire primary transcriptome has not been determined in this condition before until now. Therefore we can compare both our primary transcriptome with RNA-seq data to see which if any promoters were conditionally expressed and if they correspond to gene expression. Many of the genes that were expressed under the absence of iron had novel promoters but these mostly had weak or very weak σ^{70} motifs (Table 4.5). Iron-regulated genes may be more prone to a loose σ^{70} motif to accommodate for other regulatory factors such as the Fur binding box.

As expected, the TonB-ExbB-ExbD energy transduction system responsible for providing energy for transporting iron complexes and other substrates essential for cellular function across the outer membrane of Gram-negative bacteria, was expressed in the absence of iron (Noinaj *et al.*, 2010; Naikare *et al.*, 2013). These proteins are all encoded in one operon under two potential promoters as seen in Table 4.5. The primary promoter tss4926_Cj1628 was also identified by Handley *et al.* (2015) and has a weak σ^{70} motif, whereas the internal TSS found in tss4927_Cj1628_int is novel and overlaps the first gene *exbB2* by 1 nt, suggesting *exbB2* has the properties of a leaderless mRNA as well as possessing a canonical promoter. Leaderless mRNAs do not require a RBS or 5'UTR for translation to occur, and ribosomes solely interact with the AUG initiation start codon (Moll *et al.*, 2002; Brock *et al.*, 2008). It has been shown that an alternative formation of 61S ribosomes that lack several proteins from the 30S subunit preferentially translates leaderless mRNAs (Kaberdina *et al.*, 2009). Therefore, leaderless mRNAs may have an advantage of not requiring the 30S ribosome for recruitment of tRNAs. Other notable genes with this property include Cj1384c encoding for a hypothetical protein, which was the highest differentially expressed gene between iron-limited and iron-repleted conditions (Section 5.4.8) and this may be attributed to its leaderless property as well as having conventional translation initiation. In total, we identified 51 novel promoters that have a UTR $<$ 10 nt between the TSS and the first nucleotide of the initial codon, with 7 of these of 0 length (i.e. TSS overlaps the first nucleotide of the gene) and with a promoter motif. However, many of the identified leaderless promoters in Cappable-seq that overlapped the first nucleotide of the downstream gene did not have a promoter motif, especially if the promoter itself was not inside another gene in an operon. As seen in both tss4927_Cj1628_int and tss4149_Cj1384c_int_2 from Table 4.5, when manually searched there was a σ^{70} motif present but not within the distance restriction, and normally very close to the TSS. This could also be another property of leaderless mRNAs, although further investigation is required to confirm this.

There is a caveat to be mindful of regarding the expression levels of internal promoters, as the source of the read counts mapping to the internal promoter can either be from the activity of the internal promoter, or the expression of the gene the promoter resides in, or both. For example, *exbB1* (Cj0179) is part of a putative operon spanning genes Cj0177 to Cj0183. Cj0178 contains five internal promoters potentially driving the transcription of the downstream exbB1. The same reads will map to both the transcript driven by the promoter upstream of Cj0177 that spans the whole operon, and the transcript driven by any one of the promoters internal to Cj0178 that can drive transcription of Cj0179. Therefore, there is no way to tell if the calculated expression of the TSS transcript was due to the activity of the internal promoters alone or from the gene that is being expressed. To truly distinguish which promoters are driving genes under what conditions, Cappable-seq libraries would have to be sequenced individually for every condition. Nevertheless, this data set can still be very informative.

Another example is Cj1394, which is also in an operon. Cj1393 upstream of Cj1394 is not expressed under iron-limited conditions and did not have any significant fold-change when comparing iron limitation against iron repletion as a control. Thus, the expression (or lack of) of internal promoters upstream of Cj1394 within the gene Cj1393 can deduce which promoter is driving Cj1394. Only one of the three internal promoters within Cj1393, tss4170_Cj1393_int was also expressed in iron-limited conditions and repressed in iron-repleted conditions and therefore seem to be driving Cj1394. When the promoters were viewed on Artemis (file in supplementary materials) against the NCTC 11168 GFF file, tss4170_Cj1393_int is at the end of Cj1393 and near Cj1394, so it is likely that Cj1394 is indeed being driven by tss4170_Cj1393_int but not the other two internal promoters upstream. In another instance, as_Cj1631c_1 is preceded by an internal promoter and two antisense promoters that are within or overlap *tonB2*. Only one of the antisense promoters tss4931_as_Cj1631c were expressed in iron-repressed conditions, whereas the other two promoters and *tonB2* did not make the cutoffs. This implies that the expression of as_Cj1631c_1 is not from readthrough of *tonB2* but actually from the antisense promoter tss4931_as_Cj1631c. Without the influence of the gene upstream it is possible to infer which promoters are driving which genes under certain conditions. Users will have to take this into consideration when using this data for analysis in the future.

Fur box consensus sequences were identified in published and novel promoters revealing genes and antisense ncRNAs, which could also be iron-regulated. In the presence of iron, the holo-Fur form would be active, so it is logical for genes up-regulated and expressed in the absence of iron to contain Fur repression binding boxes so that they can be negatively regulated in the presence of iron. Many promoters associated with iron-uptake genes have a holo-Fur repression binding box (Table 4.7), which was expected since an excess of iron is toxic to cells. Both *cfrA* and *chuA* have primary promoters which agrees with this in our data, and the latter was even found as a direct target of Fur by CHiP-chip (Butcher *et al.*, 2012). Handley *et al.* (2015) also identified and found Fur/PerR binding boxes in the promoter region for *cfrA*, however, for *chuA* the primary promoter identified by Handley *et al.* (2015) is different from ours and further downstream nearer to the gene but we did not identify the same TSS. It is possible that the TSS Handley *et al.* (2015) identified for *chuA* is specific to their *fur perR* mutant whereas the promoters we identify

are associated with iron, whereas with *cfrA*, the single primary promoter is closely linked with Fur/PerR and iron. CfrA and ChuA are siderophore transporters essential for enterobactin and hemin uptake so it is no surprise that there is precise regulation of these genes by Fur. In particular, enterobactin and hemin are in the ferric form (Fe^{3+}) so if there is plenty of ferrous iron (Fe^{2+}) available and binding to Fur it would make sense to down-regulate ferric uptake systems as they are not needed (van Vliet *et al.*, 2002). In Crofts *et al.* (2018)'s study, RNA-seq data from infected human faecal samples showed that two iron acquisition pathways: *chuA*/*chuD* and *cfrA* were up-regulated >100-fold, demonstrating the important role of iron in virulence. We also show very high up-regulation of these genes in iron-limited conditions from our DESeq2 analysis. It would be interesting to see whether specific iron uptake systems for ferrous iron is up-regulated when holo-Fur regulated genes are down-regulated. The primary promoter slightly more upstream of and overlapping the secondary promoter of *chuA* that we identified contains a Fur repression box. It is possible that holo-Fur binds to the primary promoter in the presence of iron and also disrupts the secondary promoter in the process. The promoter further downstream identified by Handley *et al.* (2015) also has a Fur repression box, however, as it is difficult to distinguish between Fur and PerR binding boxes, it is also possible that the promoter is associated with PerR. The distinction between Fur and PerR binding boxes will have to considered in future analysis. The promoters for the *chuABCD* operon are divergently transcribed from Cj1613c so there may be promoter clashing. Moreover, as *chuA* is at the beginning of an operon, disruption of this gene may also affect the genes downstream. There was a discrepancy for *exbB2* as published studies report that holo-Fur activates *exbB2*, but only a holo-Fur repression sequence was found in its primary promoter.

No activation sequences were found in Table 4.5 as expected. However, two primary promoters for flagellar genes that were also identified in published studies for *flaG* and *fliI* have a predicted holo-Fur activation sequence. Butcher *et al.* (2015) found Fur and PerR regulates many flagellar biogenesis genes including *fliI*. On the other hand, a holo-Fur repression sequence was found in an antisense promoter that potentially encodes for an antisense ncRNA to *fliD*. It is possible that the antisense ncRNA represses *fliD*, and therefore holo-Fur will repress the ncRNA to allow for translation of *fliD*. Interestingly, an internal promoter upstream of *perR* has a holo-Fur activation sequence so there is a possibility that Fur contributes to the expression of PerR via this promoter. If so, a long UTR would be made which might influence the regulation of iron-regulated genes. A holo-Fur activation sequence was also found in the secondary promoter tss1788_Cj0604_2 of another global regulator Cj0604, a homologue to the polyphosphate kinase PPK2, which is involved in the production of poly-P and linked to the stringent response. Again it is very further upstream and would potentially encode for a long UTR. However, tss1788_Cj0604_2 did not have any TPM expression under any conditions. What is interesting is that *pstS* also had a secondary promoter very further upstream with a holo-Fur activation site. As the production of poly-P requires phosphates, this may not merely be a coincidence. Perhaps an excess of iron is similar to stresses which activate the stringent response. Although, strangely Cj0604 was not differentially regulated in the presence or absence of iron (Figure 5.25, Chapter 5).

There were also many hypothetical proteins that surprisingly have a holo-activation Fur sequence usually in upstream internal promoters in the preceding gene and also in antisense promoters for Cj0044c. Cj0044c has been shown to be transiently up-regulated in acid shock, and the phase-variable gene upstream Cj0045c is a suspected haemerythin as variants of this gene modulate the expression of Cj0044c (Reid *et al.*, 2008; Kim *et al.*, 2012). The potential antisense ncRNA could have some sort of repression role that is modulated by iron, although there was no differential expression of the ncRNA as_Cj0044c_2, which runs off this promoter. However, in the absence of iron at early stationary phase, Cj0044c is up-regulated by 1.25 $\log_2$ fold, suggesting it may somehow be associated with iron.

4.5.3 Evaluation of Cappable-seq data as a resource

This chapter has highlighted the vast number of novel TSS which have been analysed and categorised to reveal a more in-depth architecture of the NCTC 11168 genome. Certain regulatory elements such as leaderless mRNAs and holo-Fur binding boxes have been identified from this data set, yet there is still a plethora of transcriptional regulatory elements to be discovered and investigated. This newly re-annotated genome resulting from Cappable-seq data along with the RNAtag-Seq data from Chapter 5 will be an invaluable resource for years to come and benefit *Campylobacter* researchers greatly. Having a more robust identification of TSS as well as previously unidentified promoters for annotated orfs will aid researchers with genetic manipulation especially when creating mutant knockouts. This data set is also very useful for giving high confidence for novel putative transcripts from RNA-seq data such as ncRNAs. Newly discovered potential novel transcripts identified from this project may form the basis of future work in deciphering the molecular genetics of *C. jejuni*.

Chapter 5

The transcriptional profile of NCTC 11168 under 22 host-relevant stress conditions

5.1 Introduction

Sequencing is a method utilised to determine the nucleotide sequence of DNA or RNA. RNA-seq (also known as whole transcriptome sequencing) is the sequencing of extracted RNA from a population of cells grown under specific conditions at a certain point in time, revealing the presence and quantification of total RNA i.e. the transcriptome. RNA to be sequenced can also be enriched for either mRNA or sRNA. Advances in HTS such as RNA-seq have allowed for transcriptional profiling at the single nucleotide resolution level (Nagalakshmi *et al.*, 2008, 2010). Previous technologies such as probe-based microarrays and Sanger sequencing were originally used to study transcriptomes, but these were limited by difficulties in detecting low abundance transcripts and had lower resolution compared to RNA-seq (Nagalakshmi *et al.*, 2010).

The cost of HTS and therefore RNA-seq has decreased dramatically from two decade ago when it was first implemented, however, RNA-seq experiments can still be very expensive and time-consuming. This is especially true when scaled up to many samples as a single library is required for each sample and every condition under investigation has to be represented by replicates. rRNA depletion and library preparation kits make a significant contribution to the overall high costs of RNA-seq. RNAtag-Seq is a method created by Shishkin *et al.* (2015) which uses unique barcodes as 'tags' that are ligated to RNA samples that are then pooled together before rRNA depletion and library construction is carried out. As a result both the total number of cDNA libraries generated and the cost are significantly reduced, enabling comparative studies with a high number of samples (Shishkin *et al.*, 2015). The workflow is also simplified as the number of samples processed for rRNA depletion and library construction is significantly reduced making

large experiments easier to manage.

Many transcriptome studies have been carried out in *C. jejuni* either under standard laboratory growth conditions or focusing on a single stress condition. But it is difficult to compare the regulatory networks between these studies as different strains, media, and/or growth phases have been used. Even with the same strain, there can be minor mutations and variations from lab to lab resulting in differences in phenotype and virulence due to passaging as demonstrated in the strain NCTC 11168 (Gaynor *et al.*, 2004; Pascoe *et al.*, 2019). Nevertheless, these studies have been valuable in deciphering post-transcriptional regulatory networks and discovering novel transcripts such as ncRNAs. Transcriptomic compendiums from a collection of growth and stress conditions have been carried out in some bacterial species such as *S. enterica* serovar Typhimurium Kröger *et al.* (2013) and *S. pneumoniae* Slager *et al.* (2018) in order to build a genome-wide transcriptional network. The specific bacterial species databases generated from these transcriptomic compendiums have been an invaluable resource for their research community. To date there has not been a transcriptomic compendium conducted in *C. jejuni*, so we aim to carry out the first condition-dependent genome-wide transcriptome database of *C. jejuni* coupled with the primary transcriptome in Chapter 4 to uncover the transcriptional landscape in the *C. jejuni* reference strain NCTC 11168.

RNAtag-Seq was used in this project for sequencing of *C. jejuni* under 22 different growth and stress conditions (66 samples in total). The detailed protocol can be found in Chapter 2, Section 2.7.1, downstream analysis in Chapter 2, Section 2.5.2, and the results are presented and discussed in this chapter. *C. jejuni* is unusual in that it can infect a wide range of wild and domestic animals and humans, yet is mainly asymptomatic and acts as a commensal in avian species. Therefore, specific growth and stress conditions were selected to represent the different niches *C. jejuni* occupies and encounters in the environment. Principally, the focus was on the human and chicken host, and the transmission between these hosts. However, the suite of conditions are not a fully comprehensive list of the countless environments that *C. jejuni* may experience and are only within the scope of this project. Investigating the global transcriptional profile across many growth and stress conditions in parallel will allow a comparative study across the conditions, and reveal any crosstalk between the different conditions.

ncRNAs have been discovered in RNA-seq data serendipitously but few studies have focused on its role as a post-transcriptional regulator. ncRNAs are known to be induced by various stresses and involved in modulating stress responses (Hoe *et al.*, 2013). *C. jejuni* already lacks many global regulators present in other Gram-negative bacteria and does not encode for a homologue of the sRNA chaperone Hfq, although it does encode for a homologue of CsrA, which is a RNA-binding protein. ncRNAs could possibly play an important role in the regulation of the *C. jejuni* stress response. ncRNAs identified from Chapter 4 were analysed along with the data obtained in this chapter. The stringent response was also investigated due to the deviation from the classical stringent response in other bacteria and to deduce whether the stresses elicited a transient or long-term response for adaptation or survival.

Chapter Aims:

- Sequence individual RNA samples of NCTC 11168 from a suite of 22 growth and stress conditions using RNAtag-Seq
- Assess and perform quality control checks of RNAtag-Seq data
- Examine the transcriptional profile of NCTC 11168 for a brief overview of expression patterns across the suite of conditions
- Investigate the behaviour of stringent response genes across the suite of conditions

5.2 RNA collection and sequencing of RNAtag-Seq libraries

NCTC 11168 was grown in quadruplicate under 22 conditions listed in Table 2.1, of Section 2.1.2, Chapter 2. As mentioned previously, a replicate in each well is considered as pseudo-independent as each well undergoes the same treatment at the same time in the same environment, although cultures were inoculated from separate plates. Here we refer to these replicates as biological replicates for RNA-seq analysis. Cells were collected and stabilised with cold killing buffer to halt any transcriptional activity before long term storage at -80 °C. RNA was extracted from three of the four biological replicates in batches of 10. These were randomly selected to prevent batch effect but ensuring replicates of the same condition were in separate batches. The samples were DNase treated and quality checked before proceeding with RNAtag-Seq and sequencing. A 2 nM library pool of 10 conditions in triplicate (30 samples) of NCTC 11168 were initially sequenced on half a 150 cycle v3 MiSeq cartridge on the Illumina MiSeq platform (see Chapter 2, Section 2.6 for the full protocol). This was done to quality check the libraries and samples and ensure a working protocol before proceeding with sequencing on the Illumina NextSeq.

Normally the MiSeq automatically converts binary base call (BCL) files into fastq files. However, this was not possible with RNAtag-Seq libraries as it was discovered that the i5 tag was incorporated in the read rather than the index sequence region. Thus, BCL files were converted to fastq files using the bcl2fastq software (Illumina) according to the i7 index, and then successfully de-multiplexed by executing Example_tile.py (written by Leonidas Souliotis) according to the tag sequence (see Chapter 2, Section 2.7.1 for details). Although the data was limited, with only $\approx$2 fold-coverage for each of the 66 samples (see Section 5.3), our preliminary analysis confirmed that all samples were represented in our pool and the tagging had worked as expected showing proof of principle. However, as the data was insufficient for further analysis, the same libraries were subsequently sequenced again on the Illumina NextSeq platform with a NextSeq 550 high output cartridge 150 cycles for deeper coverage.

A further 12 conditions (and some repeats of the previous 10 conditions) underwent the same procedure from growth curves to RNA extraction to library preparation. The samples were successfully sequenced on the Illumina NextSeq on a separate NextSeq 550 high output cartridge 150 cycles and de-mulitplexed as before giving a total of 22 conditions altogether with three biological replicates, i.e. 66 fastq files for each sample.

5.3 Data analysis: quality control

5.3.1 Mapping of RNAtag-Seq data

Reads from the fastq files were mapped to NCTC 11168 genome from NCBI Accession number: AL111168 AL139074-AL139079. Table 5.1 shows that all samples had over 90 % alignment rate from the initial MiSeq run. One of the replicates of anaerobic stress had the lowest alignment rate of 90.56 %. The unmapped reads of this sample were filtered using SAMtools and put through BLAST (Basic local alignment search tool) revealing alignment to the kitome. There was no dramatic difference between alignment rates from the conditions sequenced on the MiSeq and the NextSeq platforms (Figure 5.1). For most of the samples from the NextSeq run, there was also over 90 % alignment rate apart from the cold-related stresses: cold stress, 5 % chicken juice, and 100 % chicken juice. Especially for 100 % chicken juice the alignment rates were very low. As chicken juice contains RNA from chicken, the unmapped reads for 100 % chicken juice and 5 % chicken juice samples were subsequently mapped against the chicken genome *Gallus gallus* GRCg6a taken from NCBI to confirm this. The alignment rates are shown in Table 5.2 revealing a large percentage of reads aligning to the chicken genome in 100 % chicken juice, and a small percentage in 5 % chicken juice which when added to the reads aligning to NCTC 11168 comprise over 90 % alignment. Consequently, only the mapped reads filtered from BAM files were taken forward for coverageBed to generate read counts and further analysis. The GFF file obtained from NCBI Accession number: AL111168 AL139074-AL139079 and corresponding read counts from coverageBed were used as inputs for DESeq2 analysis.

Condition	MiSeq total number of reads (alignment rate %)	Nextseq total number of reads (alignment rate %)
37 °C exponential	-	6,686,795 (97.73)
	-	6,639,898 (97.95)
	-	6,343,023 (96.79)
37 °C early stationary	48,521 (96.57)	997,625 (96.66)
	173,737 (96.68)	3,617,569 (98.47)
	199,094 (98.99)	4,002,662 (98.74)
37 °C late stationary	-	5,860,986 (98.36)
	-	9,291,862 (98.20)
	-	5,317,738 (98.30)
42 °C exponential	177,785 (97.73)	3,678,800 (97.72)
	245,664 (98.71)	5,103,364 (98.49)
	249,834 (97.61)	5,053,884 (97.68)
42 °C early stationary	226,217 (97.64)	4,526,312 (97.89)
	254,146 (97.91)	5,135,949 (98.02)
	118,013 (98.16)	2,316,623 (98.17)
42 °C late stationary	82,174 (97.77)	1,795,690 (97.90)
	78,607 (97.23)	1,719,313 (97.49)
	62,814 (97.27)	1,339,577 (97.49)
anaerobic	116,103 (90.56)	2,380,981 (92.54)
	152,820 (97.16)	3,224,401 (97.33)
	168,948 (96.79)	3,453,143 (97.07)
heat	141,717 (96.92)	3,829,883 (97.04)
	350,198 (94.10)	9,489,793 (94.98)
	133,748 (97.52)	3,585,122 (97.46)
iron limitation exponential	-	4,132,251 (98.42)
	-	7,036,325 (98.44)
	-	9,772,440 (98.38)
iron limitation early stationary	-	2,794,776 (98.33)
	-	4,144,108 (95.95)
	-	4,135,483 (98.13)
iron repletion exponential	-	1,849,794 (98.61)
	-	1,061,051 (98.40)
	-	2,376,302 (98.50)
iron repletion early stationary	-	12,169,415 (98.02)
	-	7,612,421 (96.11)
	-	8,771,920 (98.25)

Table 5.1 continued from previous page

Condition	MiSeq total number of reads (alignment rate %)	Nextseq total number of reads (alignment rate %)
sodium deoxycholate exponential	-	9,064,148 (97.28)
	-	14,482,733 (97.98)
	-	17,365,335 (98.00)
sodium deoxycholate early stationary	51,413 (96.85)	1,136,371 (97.24)
	93,650 (97.71)	2,086,175 (97.86)
	81,743 (99.14)	1,787,877 (98.84)
nutrient starvation	83,694 (98.77)	1,849,794 (98.61)
	47,974 (98.52)	1,061,051 (98.40)
	111,325 (99.16)	2,376,302 (98.50)
hyperosmotic stress	-	7,924,665 (96.15)
	-	5,877,781 (92.35)
	-	5,648,810 (96.13)
Acid shock	-	5,006,590 (97.70)
	-	3,404,640 (95.24)
	-	2,617,357 (97.40)
peroxide	-	5,409,400 (96.37)
	-	4,269,498 (86.91)
	-	4,656,025 (91.89)
cold	-	7,368,505 (74.61)
	-	4,301,028 (68.87)
	-	3,728,232 (80.97)
5 % chicken juice	-	4,473,162 (80.46)
	-	7,445,715 (63.13)
	-	7,466,358 (77.15)
100 % chicken juice	-	331,286 (7.64)
	-	294,908 (2.61)
	-	222,172 (1.40)
GSNO	-	6,872,870 (95.78)
	-	3,359,424 (94.56)
	-	10,012,137 (96.63)

Table 5.1: Table of the total number of reads and percentage alignment rate for each biological replicate of NCTC 11168 from the MiSeq and NextSeq output of 22 RNAtag-Seq libraries. Only 10 conditions were tested on the MiSeq. Two conditions were sequenced twice on the NextSeq but the duplicates are not included in this table.

Sample	Total number of reads	Overall alignment rate
100 % chicken juice	331,286	70.88 %
	294,908	46.40 %
	222,172	65.98 %
5 % chicken juice	4,473,162	10.98 %
	7,445,715	23.37 %
	7,466,358	14.33 %

Table 5.2: Table of the total number of reads and percentage alignment to the chicken genome.

5.3.2 DESeq2 analysis and quality control

The initial 10 conditions that were sequenced on the Nextseq platform samples were normalised together using DESeq2. Regularized log (rlog) transformation was used for a Principal Component Analysis (PCA) of the raw data and visually represented in Figure 5.1, using the option `blind = TRUE`. All three biological replicates from each condition cluster well together, apart from one outlier from 37 °C exponential phase (circled in dark green), therefore 37 °C was repeated in the next run. Replicates circled in red representing samples for the iron repletion early stationary condition were also repeated as it was suspected that the bottle of MEMα used for the first preparation to implement the stress was not at the correct pH. Although Figure 5.1 only includes 10 of the 22 conditions, it was reassuring to see that the nutrient deprived conditions: starvation, 42 °C early and late stationary, and 37 °C early stationary were close together.

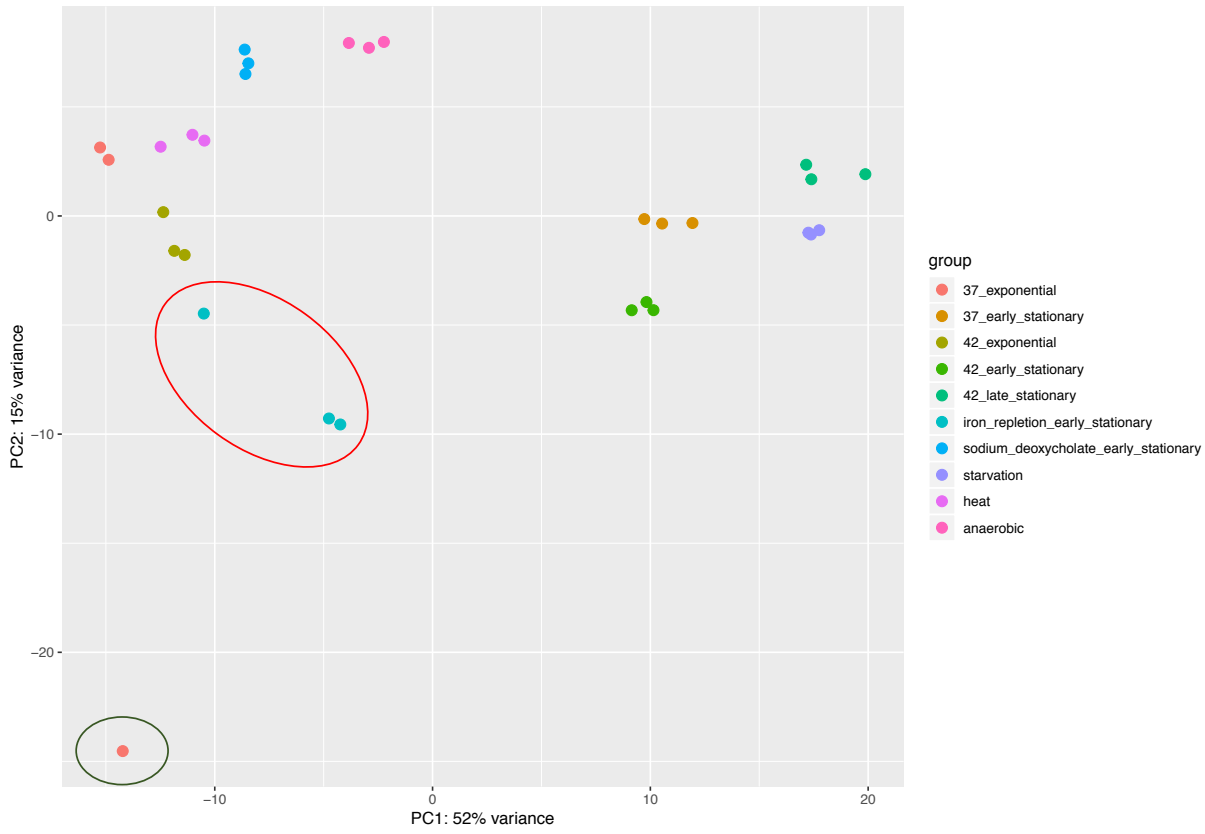


Figure 5.1: PCA plot of transformed raw expression data of 10 conditions sequenced from RNAtag-Seq libraries. An outlier from 37 °C exponential phase is circled in dark green (bottom left) and all iron repletion early stationary replicates are circled in red.

After the second Nextseq run, data from the additional 12 conditions and any repeats were analysed together with the previous 10 conditions giving a total of 22 conditions. A PCA of the repeated conditions (37 °C exponential and iron repletion early stationary) confirmed a better reproducibility between the biological replicates (Figures 5.2 and 5.3). Biological replicates clustered together by Principal component 2 (PC2) in Figure 5.2 and Principal component 1 (PC1) in Figure 5.3 indicating also that the biological variation observed was not due to batch variation and that RNAtag-Seq is reproducible. The outlier for 37 °C exponential phase in Figure 5.2, as was seen in Figure 5.1, has 88 % variation from the other replicates, thus only the three biological replicates from the second run of 37 °C exponential phase were subsequently used for further analysis. Only the three iron repletion early stationary replicates from the second run were also used for further analysis since they cluster together more closely.

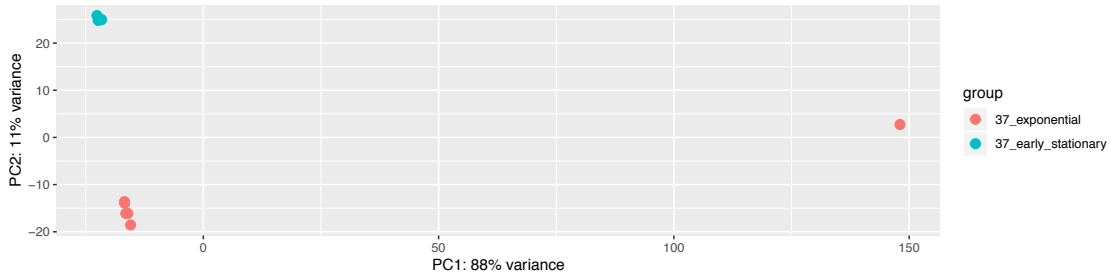


Figure 5.2: PCA plot of plot of 37 °C early stationary against 37 °C exponential phase from two different RNAtag-Seq library sequencing runs. Samples from both runs cluster together apart from the outlier in 37 °C exponential.

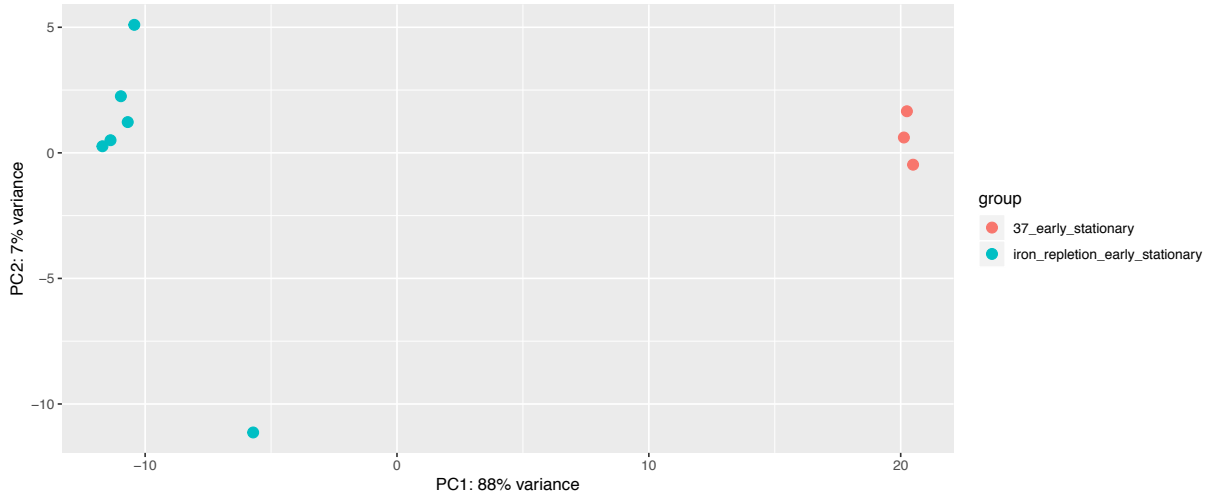


Figure 5.3: PCA plot of plot of 37 °C early stationary phase against iron repletion early stationary phase from two different RNAtag-Seq library sequencing runs. Samples from both runs cluster together.

Prefix	Meaning
G	Growth-related
B	Bile-related
I	Iron-related
C	Chicken juice and cold-related
S	Stress-related

Short-hand	Meaning
E	exponential
S	early stationary
L	late stationary
a	all
b	sodium deoxycholate
IL	iron limitation
IR	iron repletion
cj100	100 % chicken juice
cj5	5 % chicken juice
37	37 °C
42	42 °C
4	4 °C
_(n)	replicate

Table 5.5: Key for condition and comparison names used for plots and heat maps. (n) indicates any number.

Before running DESeq2 on all 66 samples (of the 22 conditions in triplicate), published ncRNAs and putative novel ncRNAs identified in Chapter 4 were added to the GFF file taken from NCBI Accession number: AL111168 AL139074-AL139079. A PCA plot of all conditions (Figure 5.4) without the ncRNAs added with the key for condition names in Table 5.5 shows the biological replicates cluster very well together which indicates there is not much variation between replicates. The maximum variation depicted by PC1 seems to be explained by the difference in growth phase/nutrient depletion as the conditions clustering on the left circled in red are all nutrient deprived conditions. This includes 37 °c early stationary, 37 °C late stationary, 42 °C early stationary, 42 °C late stationary, starvation, and iron limitation early stationary. However, there is another cluster circled in dark green at the bottom with osmotic stress, 100 % chicken juice, and iron repletion early stationary.

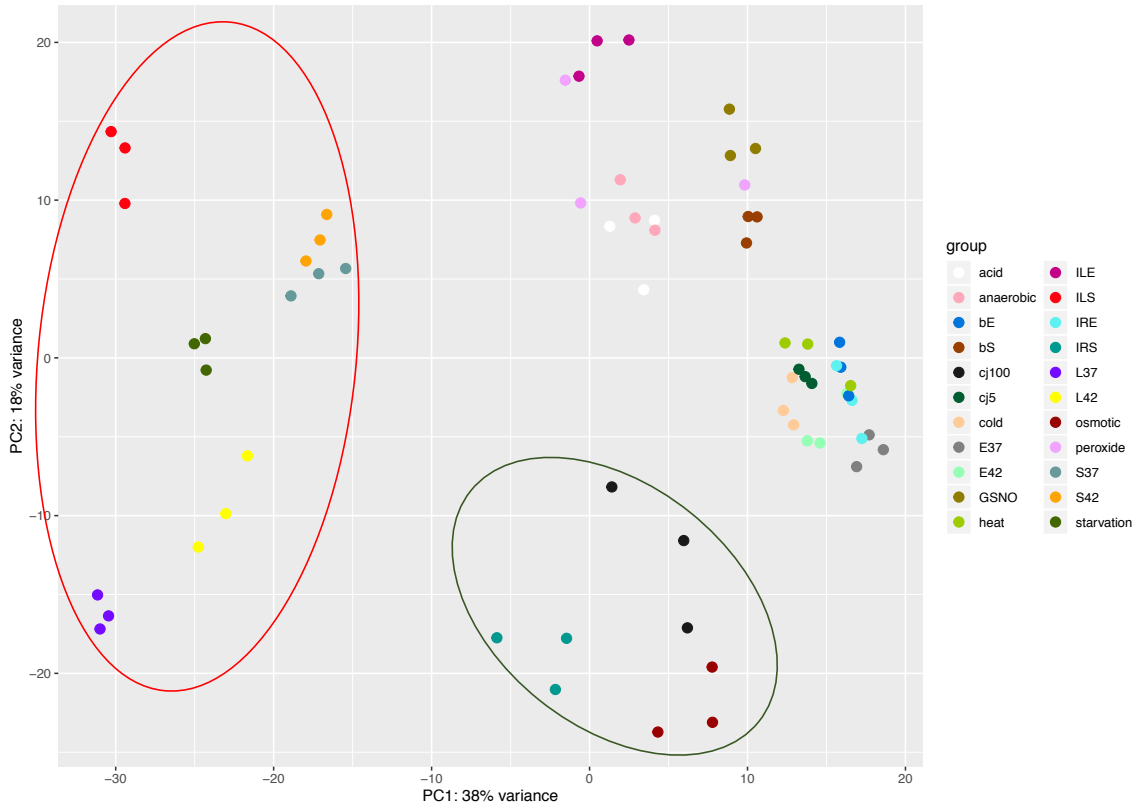


Figure 5.4: PCA plot of transformed raw expression data from all 22 conditions. Conditions clustering on the left (circled in red) are nutrient deprived conditions, whereas the cluster at the bottom (circled in dark green) are iron repletion early stationary (IRS), 100 % chicken juice (cj100), and osmotic stress.

However, when read counts from published and putative novel ncRNAs (obtained from Chapter 4) were added, the PCA profile changes (Figure 5.5). In particular, 100 % chicken juice replicates are not closely clustered together anymore and other conditions such as peroxide stress replicates have also slightly moved. Overall, replicates in the other conditions are still clustered together and the variance of PC1 is reduced to 32 %. The read count matrix used for DESeq2 analysis containing the coverage read count of ncRNAs were subsequently used for all analysis.

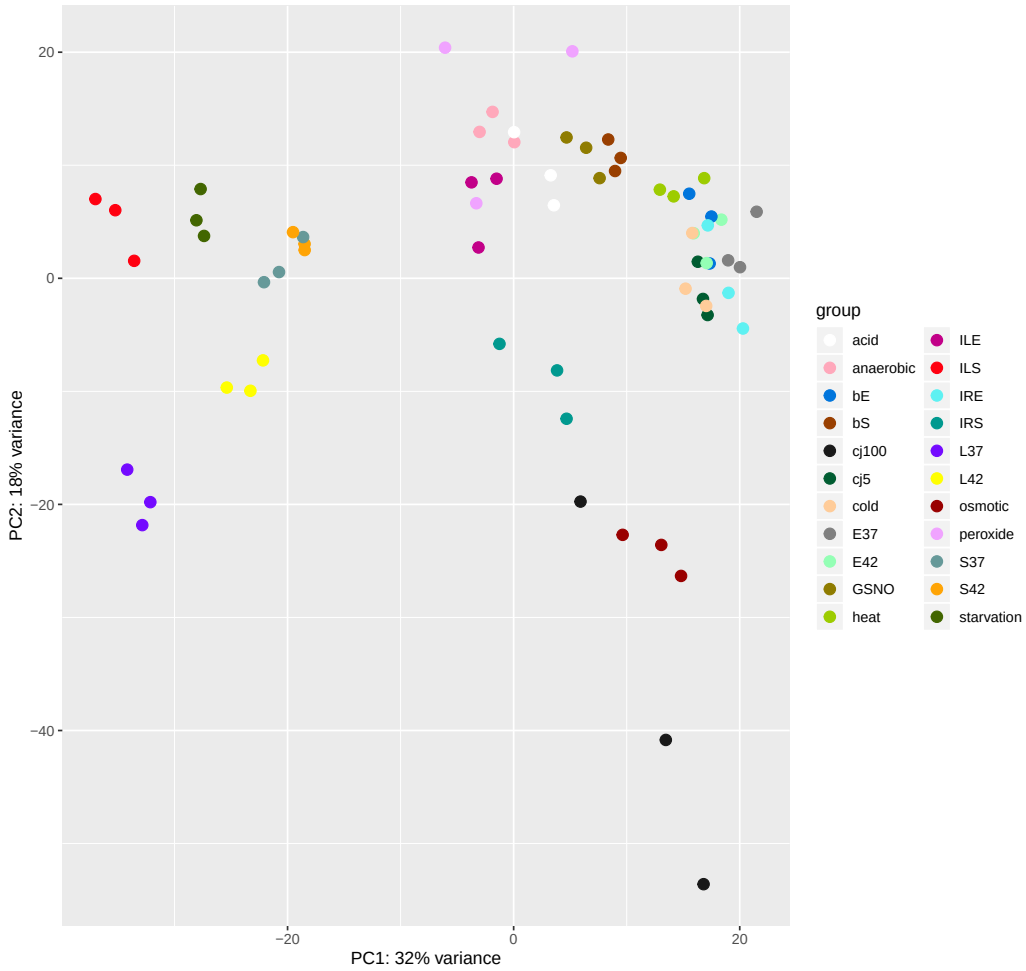


Figure 5.5: PCA plot of transformed raw expression data including ncRNAs from all 22 conditions. After the addition of ncRNAs the profiles of some conditions changed.

When the ncRNAs were analysed exclusively in Figure 5.6, again the PCA profile changes. The variation in PC1 is reduced but for some conditions, namely 100 % chicken juice (cj100), the replicates are no longer clustered as closely together in line with Figure 5.5. However, other conditions such as acid shock show the replicates are further apart suggesting the profile of ncRNAs is more variable between replicates than gene expression. There is also no longer an obvious cluster/distinction with nutrient-deprived conditions.

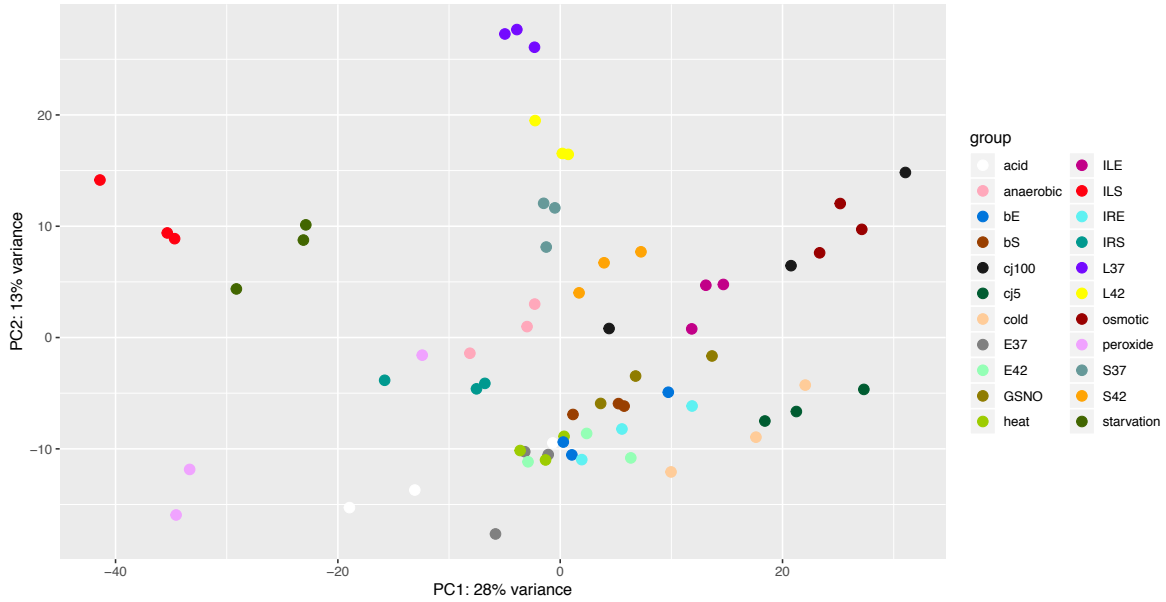


Figure 5.6: PCA plot of transformed raw expression data of ncRNAs only from all 22 conditions.

The rlog transformed data for all 66 samples were clustered using complete Euclidean distance hierarchical clustering to examine the similarities between replicates and conditions (Figure 5.7). All replicates from each condition clustered together with the exception of one replicate each from cold and cj5 conditions which clustered with the other two replicates of the other condition, but both of these conditions are also clustered together. The distance between replicates was very similar apart from cj100, although nuances may have been missed due to the high variation seen in cj100 replicates possibly due to the added ncRNAs and the low coverage across the genome with other conditions as was seen in the PCA plot (Figure 5.4).

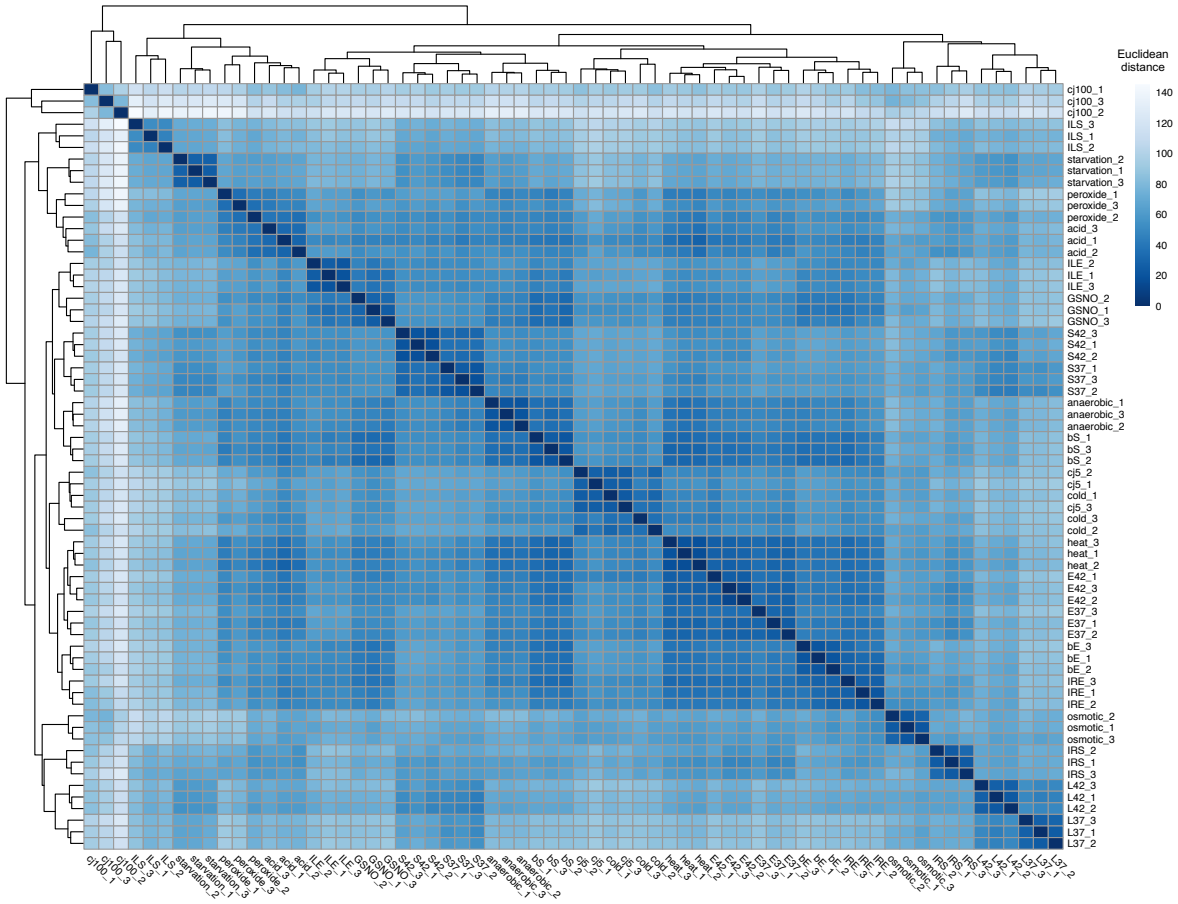


Figure 5.7: Heat map of a distance matrix of rlog transformed expression values including ncRNAs of all samples. Complete Euclidean hierarchical clustering was used. Samples from each condition are shown on the x and y-axis with replicates indicated by the number preceded by _.

When the ncRNAs were removed from the distance matrix (Figure 5.8), again one replicate from cj5 and cold stress had the same clustering as before but other conditions were clustered differently together. There is a more evident group of conditions at the bottom comprising of: ILS, L42, L37, starvation, S42, and S37 which are more dissimilar to the other conditions but are fairly similar to each other. These conditions seem to be related to nutrient deprivation. In addition, osmotic stress clusters with 100 % chicken juice again as was seen in the PCA plot (Figure 5.4).

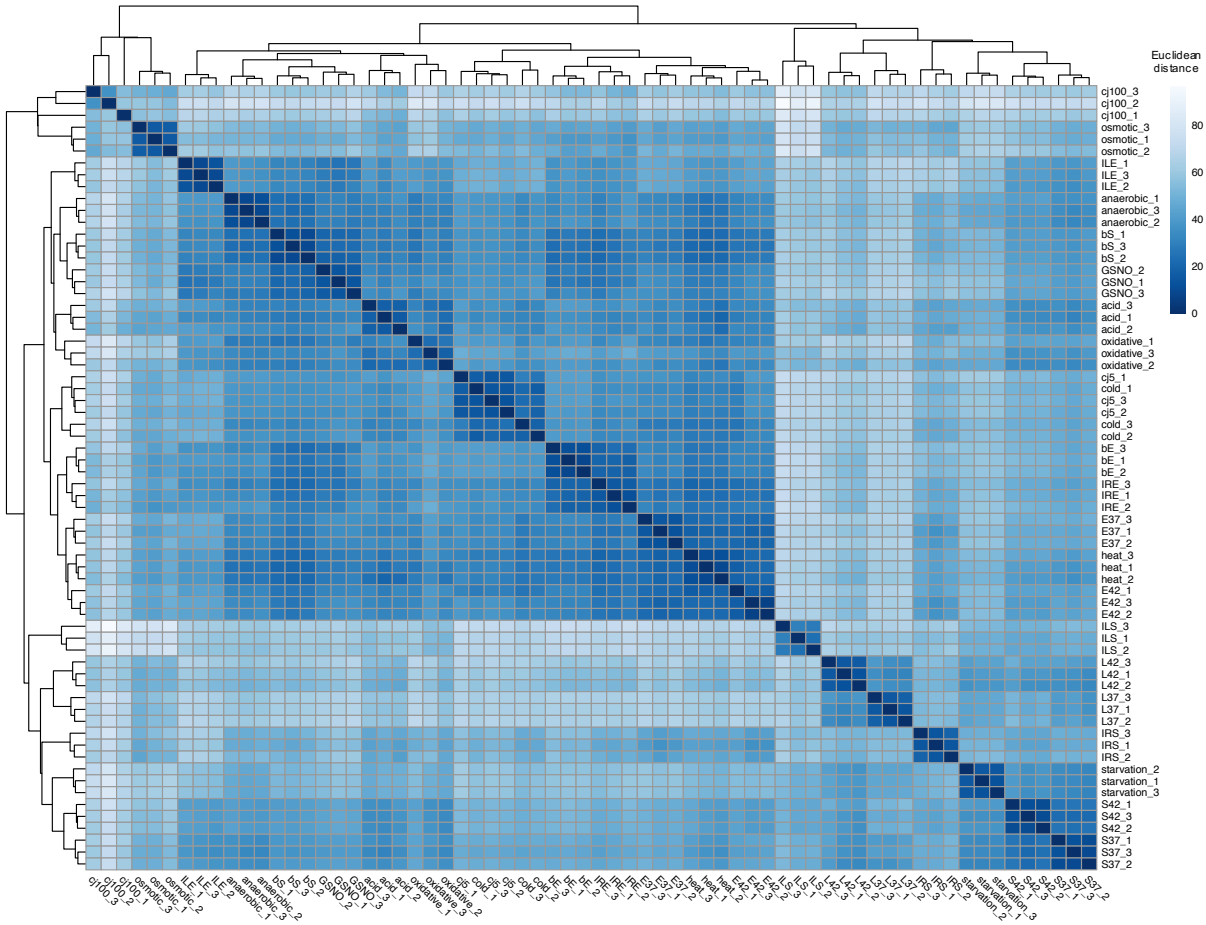


Figure 5.8: Heat map of a distance matrix of rlog transformed expression values without addition of ncRNAs of all samples. Complete Euclidean hierarchical clustering was used. Samples from each condition are shown on the x and y-axis with replicates indicated by the number preceded by _.

Figure 5.9 is a euclidean distance heat map of ncRNAs only, which further emphasises the variability in cj100 replicates. The clustering is different to the previous heat maps, implying the profile of ncRNAs vary greatly from the profiles of genes. One replicate from acid, peroxide, heat, cold, 42 °C early stationary phase, and nitrosative stress (GSNO) do not cluster with their other replicates. ILS, starvation, IRS, two replicates of peroxide stress and acid shock, and one replicate of E37 clustered together, whereas cj100 replicates are a separate cluster but clusters more closely with the rest of the conditions.

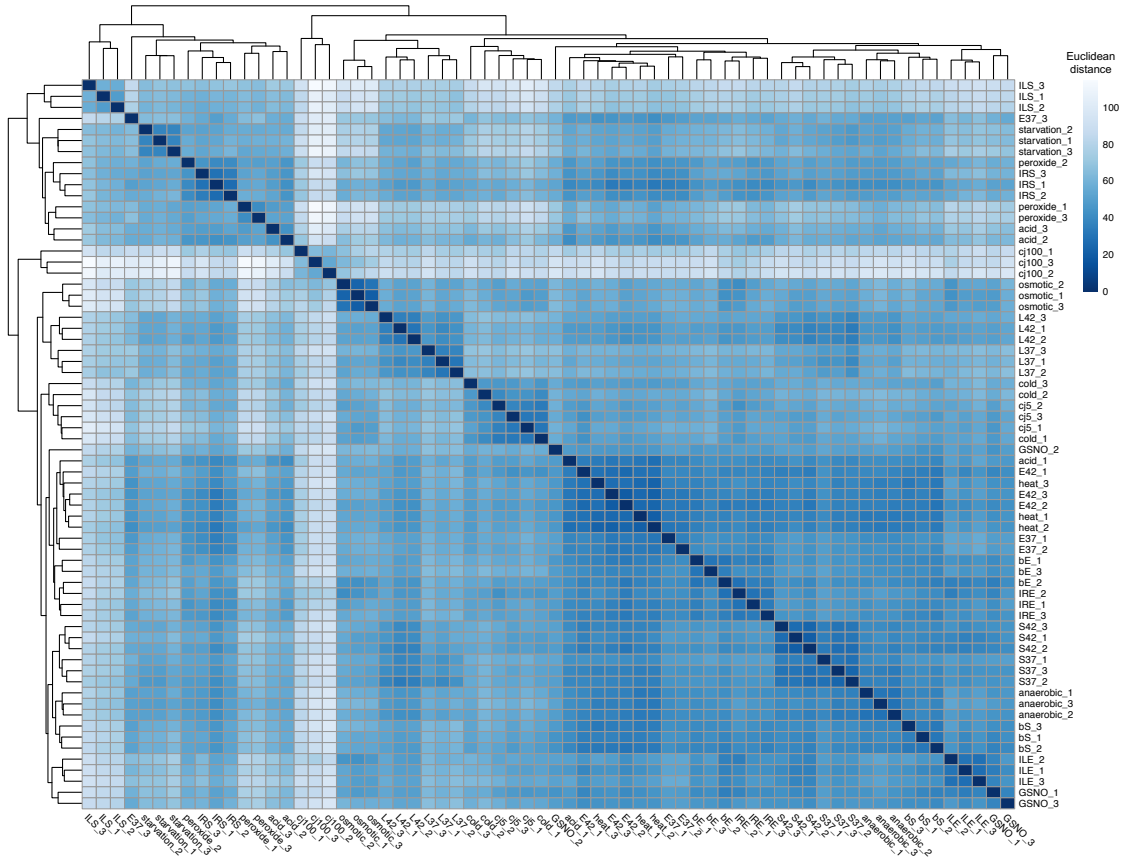


Figure 5.9: Heat map of a distance matrix of rlog transformed expression values with ncRNAs only of all samples. Complete Euclidean hierarchical clustering was used. Samples from each condition are shown on the x and y-axis with replicates indicated by the number preceded by _.

5.4 Functional pathway enrichment analysis of single and grouped comparisons using STRING

The stress and growth conditions in this project were designed to focus on changes in transcriptional expression due to a single environmental condition at a time. As we have a compendium of many growth and stress conditions, it is possible to group some of the samples together so that we can ignore the effects of changes due to multiple variables, which may give additional information when interpreting the data. Therefore, single pairwise comparisons and grouped comparisons (Table 5.6) were both used to make a compiled table of $\log_2$ fold-change values from DESeq2 (available at Git Hub[1]) for functional enrichment analysis. The values were filtered with *p*-adjusted value ≤ 0.05 and $\log_2$ fold-change cut-offs > 1 and < -1 so that only highly significantly changing genes were retained. Significantly down-regulated genes ($\log_2$ fold-change < -1) were assigned as 0, whereas significantly up-regulated genes were assigned a value of 1. For each comparison, the significantly changed genes were run through the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) enrichment analysis performed on R using the package STRINGdb (Szklarczyk *et al.*, 2015). ncRNAs were not included in this analysis as the STRING database only contains predicted, published, and/or experimentally validated gene and protein associations, but putative ncRNAs have not yet been officially annotated.

Pathways from the KEGG database were used for the enrichment analysis. The pathways were also filtered by *p*-adjusted value ≤ 0.05 (using Benjamini and Hochberg false discovery rate) for all comparisons so that the pathway was only retained if it is statistically significantly enriched in at least one comparison. The *p*-adjusted values were transformed by $-\log_{10}$ (inverse) for the up-regulated genes and $\log_{10}$ for the down-regulated genes for ease of visualisation in the following heat maps Figures 5.10-5.19. Values were combined together and plotted using complete hierarchical clustering with Pearson correlation for the pathways and Euclidean distance for all pairwise and grouped comparisons in Table 5.6, which also show the corresponding column names for each comparison on the heat maps. The prefix character of each column name indicates which group the comparison belongs to: G for growth and temperature, B for bile, I for iron-related comparisons, C for cold and chicken juice, and S for stress. Some pathways had very strong $\pm\log_{10}$ *p*-adjusted values so these were scaled accordingly to a maximum and minimum of 10 and -10. The genes from each comparison were compared with the KEGG database[2] to see what genes enriched for the pathways in the heat maps. As the option of IEA (inferred from electronic annotation) was used for the enrichment analysis, there were some cases where pathways were enriched from STRING but did not appear on the KEGG website for *C. jejuni*. There was also redundancy in some pathways as both general and specific pathways were included. Only the KEGG pathways which appeared on the KEGG database for specific pathways were discussed in detail.

[1] `https://github.com/Jenna-Lam/Supplementary-material-for-thesis`
[2] `https://www.genome.jp/kegg/tool/map_pathway2.html`

Groups	Column name	Control	Comparison
Growth phase and temperature effect	G-S/E-37	37 °C exponential	37 °C early stationary
	G-L/E-37	37 °C exponential	37 °C late stationary
	G-L/S-37	37 °C early stationary	37 °C late stationary
	G-S/E-42	42 °C exponential	42 °C early stationary
	G-L/E-42	42 °C exponential	42 °C late stationary
	G-L/S-42	42 °C early stationary	42 °C late stationary
	G-42/37-E	37 °C exponential	42 °C exponential
	G-42/37-S	37 °C early stationary	42 °C early stationary
	G-42/37-L	37 °C late stationary	42 °C late stationary
	G-S/E-a	37+42 °C exponential	37+42 °C early stationary
	G-L/E-a	37+42 °C exponential	37+42 °C late stationary
	G-L/S-a	37+42 °C early stationary	37+42 °C late stationary
	G-42+37-a	all 37 °C	all 42 °C
Effect of bile	B-b/37-E	37 °C exponential	sodium deoxycholate exponential
	B-b/37-S	37 °C early stationary	sodium deoxycholate early stationary
	B-S/E-b	sodium deoxycholate exponential	sodium deoxycholate early stationary
	B-S/E-a	37 °C+sodium deoxycholate exponential	37 °C+sodium deoxycholate early stationary
	G-S/E-37	37 °C exponential	37 °C early stationary

Table 5.6 continued from previous page

Groups	Column name	Control	Comparison
Effect of iron limited and iron repleted conditions	I-IR/IL-E	iron limitation exponential	iron repletion exponential
	I-IR/IL-S	iron limitation early stationary	iron repletion early stationary
	I-S/E-IL	iron limitation exponential	iron limitation early stationary
	I-S/E-IR	iron repletion exponential	iron repletion early stationary
	I-37/IL-E	iron limitation exponential	37 °C exponential
	I-37/IL-S	iron limitation early stationary	37 °C early stationary
	I-37/IR-E	iron repletion exponential	37 °C exponential
	I-37/IR-S	iron repletion early stationary	37 °C early stationary
	I-IR/IL-a	all iron limitation	all iron repletion
	I-37/IL-a	all iron limitation	both phases of 37 °C
	I-37/IR-a	all iron repletion	both phases of 37 °C
Cold and chicken juice effect	C-4/E37	37 °C exponential	cold (4 °C)
	C-cj5/E37	37 °C exponential	5 % chicken juice
	C-cj100/E37	37 °C exponential	100 % chicken juice
	C-cj5/4	cold (4 °C)	5 % chicken juice
	C-cj100/4	cold (4 °C)	100 % chicken juice
	C-cj100/cj5	5 % chicken juice	100 % chicken juice
	C-a4/37	37 °C exponential	all chicken juice and cold (4 °C)
	C-acj/E37	37 °C exponential	all chicken juice

Table 5.6 continued from previous page

Groups	Column name	Control	Comparison
Effect of various stresses	S-acid/E37	37 °C exponential	acid
	S-anaerobic/E37	37 °C exponential	anaerobic
	S-GSNO/E37	37 °C exponential	nitrosative
	S-heat/E37	37 °C exponential	55 °C
	S-osmotic/E37	37 °C exponential	osmotic
	S-peroxide/E37	37 °C exponential	peroxide
	S-starvation/S37	37 °C exponential	nutrient starvation
	S-stress-a/E37	37 °C exponential	all stresses

Table 5.6: Table of pairwise and grouped comparisons used for DESeq2 analysis. Comparisons are also categorised in groups for functional pathway enrichment analysis in STRING.

5.4.1 Overview of comparisons

A heat map of all 44 unique comparisons from Table 5.6 with KEGG pathways that are significantly activated (derived from the up-regulated genes in red) and inactivated (derived from the down-regulated genes in blue) is shown in Figure 5.10. Pathways which were not significantly enriched (with $\log_{10}$ p-adjusted value ≤ 0.05) are in white. There were instances where pathways such as ABC transporters were enriched by both up-regulated (red) and down-regulated (blue) genes as highlighted by the green box. This is because some pathways have sets of genes that are involved in diverse pathway networks, so one set may be up-regulated whilst another down-regulated. Otherwise normally there would be either activation or inactivation of a pathway per comparison. Figure 5.10 has two major column clusters at the top dendrogram that are sub-divided into annotated coloured sub-clusters denoted by the 'comparisons' key.

The main cluster on the left consists of the sub-clusters 4, 6, and 2 whereas the right main cluster has the sub-clusters 1, 5, and 3. The left main cluster is separated from the right cluster by the down-regulation of genes involved in a small cluster of pathways that are involved in Global and overview maps branch of Metabolism from KEGG highlighted by the orange box. Within this cluster, sub-cluster 4 is clustered separately from sub-clusters 6 and 2, as the latter two have more similar enriched pathways. Sub-cluster 4 contains all comparisons that have cj100 (100% chicken juice) as the "treatment", defined by many inactivated pathways and very few activated pathways in some of the other comparisons. Sub-cluster 6 contains all the late stationary comparisons against exponential phase at both host temperatures but also G-L/S-42 (late stationary against early stationary at 42 °C) and surprisingly osmotic stress compared against 37 °C exponential (S-osmotic/E37). Osmotic stress clusters separately from the other stresses (which are all in the main cluster on the right) an observation that is in agreement with the PCA plot in the previous section (Figure 5.1) due to the similarities in the pathway response patterns to the growth phase comparisons.

Sub-cluster 2 contains the remaining growth-phase dependent comparisons, which are all early stationary against exponential comparisons in both host temperatures but also late stationary against early stationary at 37 °C (G-L/S-37) and the grouped comparison independent of temperature (G-L/S-a), and strangely B-b/37-E and I-37-IL-S, which are bile against no bile exponential phase, and MH2 broth against MEMα at early stationary respectively. The defining characteristic of sub-cluster 2 and 6 is the activation of the Flagellar assembly pathway, which is known to be growth-phase dependent, and activation of the Ribosome pathway in a few comparisons as highlighted by the light blue box. Flagellar assembly was only enriched in the growth-phase related comparisons, whereas the Ribosome pathway was not enriched in S-osmotic/E37, G-S/E-a, and G-S/E-42.

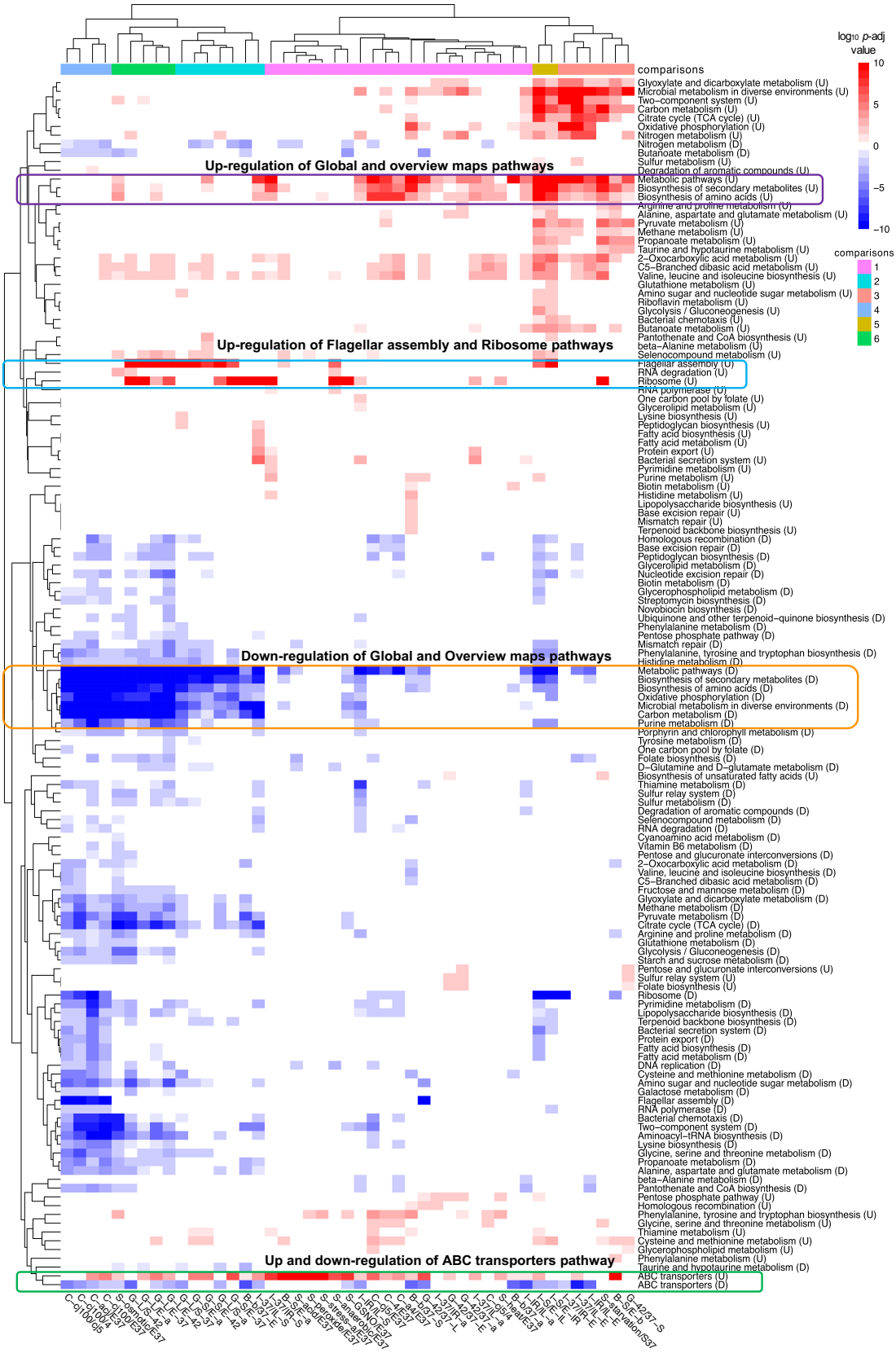


Figure 5.10: Heat map of functional enrichment analysis of all 44 comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes. Scale from red to blue indicates the intensity of the $\log_{10}$ *p*-adjusted value (inverse for up-regulated genes) and the comparisons legend indicates the sub-clusters.

The main cluster on the right has sub-clusters 1, 5, and 3. Sub-cluster 5 has two comparisons I-S/E-IL and I-S/E-IR, which are both early stationary against exponential phase in the absence and presence of iron respectively in MEMα. There are a few enriched pathways just for this sub-cluster only including activation of Amino sugar and nucleotide sugar metabolism, Riboflavin metabolism, and Glycolysis/Gluconeogenesis and many other pathways which have a similar enrichment pattern. Flagellar assembly is also activated in this sub-cluster. Sub-cluster 5 clusters with sub-cluster 3, which includes the iron-related comparisons I-37/IR-E, I-37/IL-E, and I-IR/IL-E; S-starvation/S37; B-S/E-b; and G-42/37-S. Both sub-clusters are defined by the activation of pathways under the Global and overview maps branch of Metabolism from KEGG highlighted by the purple box. Sub-cluster 1 contains the remaining comparisons populated by all the stress comparisons other than S-osmotic/E37, the remaining cold stress conditions, and remaining iron and bile conditions. There are not many clear patterns in sub-cluster 1 though most of the stress comparisons cluster together apart from S-heat/E37 along with B-S/E-a and I-37/IR-S which all activate ABC transporters strongly. For a better insight on the patterns observed, all comparisons were categorised into groups as shown in Table 5.6 and analysed further.

5.4.2 Growth phase and temperature effect

Growth phase and temperature were compared together as samples from both 37 °C and 42 °C were taken at three different growth phases (Figure 5.11). As *C. jejuni* is able to colonise chickens as a commensal but cause symptomatic infection in mammals including humans the difference in host body temperature may induce different transcriptional responses. Here we can see the comparisons are split into two main clusters as demonstrated by the top dendrogram and sub-divided into three sub-clusters as seen in Figure 5.11. The left cluster is sub-cluster 1, which consists of all the temperature comparisons and the right cluster is sub-clusters 2 and 3, which are all the growth phase dependent comparisons. The main differences are due to the down-regulation of many general metabolic pathways under Global and overview maps as highlighted by the orange box and the strong activation of the Flagellar assembly pathway highlighted by the green box. There is also sporadically enriched Ribosome and ABC transporters from up-regulated genes. Sub-cluster 3 has all the late stationary against exponential comparisons for both host temperatures and also G-L/S-42 late stationary against early stationary at 42 °C. Whereas sub-cluster 2 has all the early stationary against late stationary comparisons and the remaining late stationary against early stationary comparisons. As G-L/S-42 clusters separately from G-L/S-37 this indicates there are differences in the transition from early stationary phase to late stationary phase between the host temperatures.

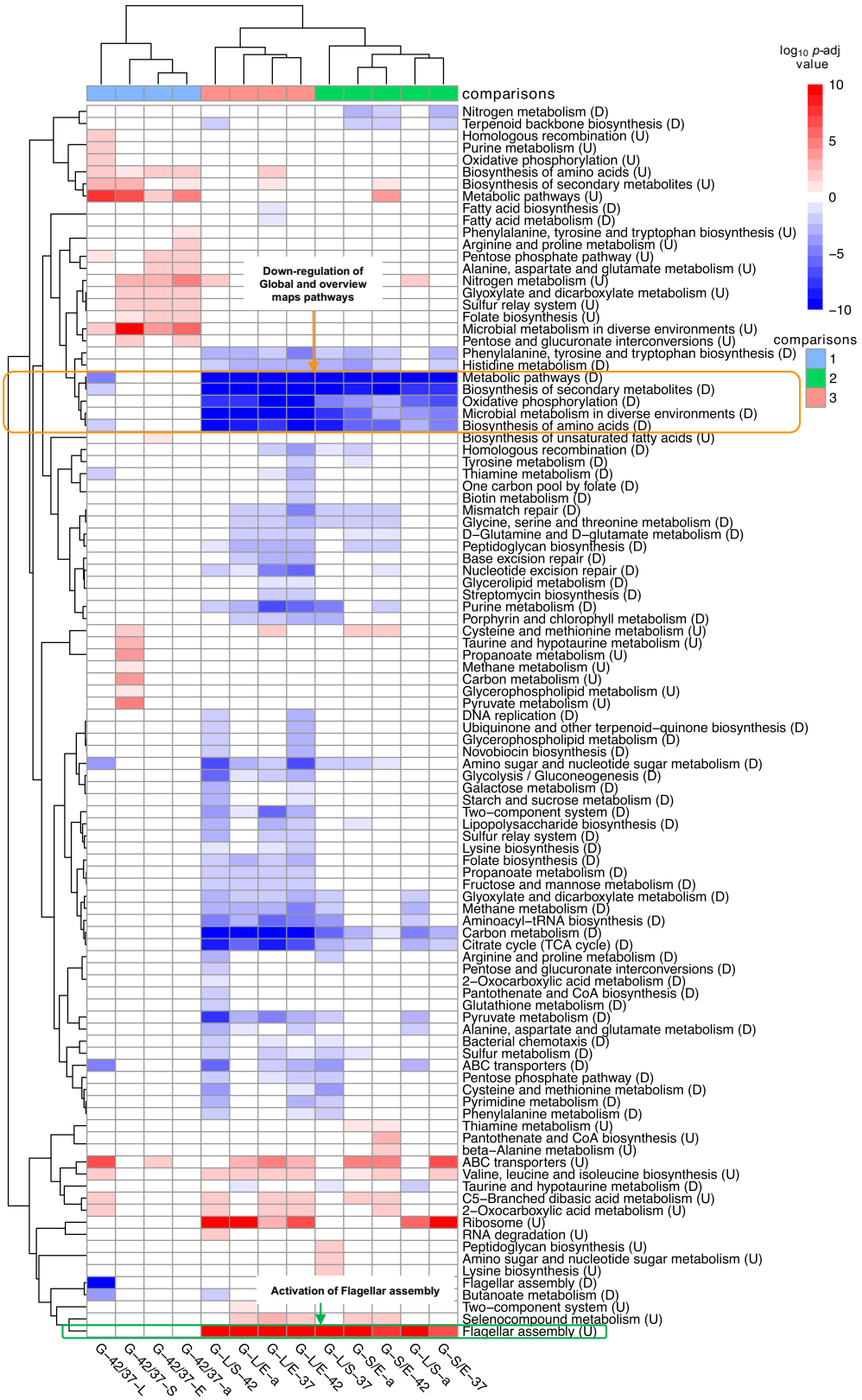


Figure 5.11: Heat map of functional enrichment KEGG pathway analysis of growth and temperature related comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes. Scale from red to blue indicates the intensity of the $\log_{10}$ *p*-adjusted value (inverse for up-regulated genes) and the comparisons legend indicates the subclusters.

Flagellar assembly was enriched by up-regulated genes in sub-clusters 2 and 3 across both host temperatures but not in sub-cluster 1 between the host temperature comparisons. However, Flagellar assembly was strongly inactivated in G-42/37-L by the genes shown in Figure 5.12 taken from the KEGG pathway database[3] so there were differences in the expression of late genes between the host temperatures. Many genes that were down-regulated were late gene products and the transcriptional regulator σ^{28} (also known as FliA). For all the comparisons involving early stationary against exponential phase in both host temperatures and combined the same up-regulated genes enriched for this pathway (Figure 5.13). On the other hand, the up-regulated genes enriching for all the late stationary against early stationary or exponential phases shown in Figure 5.14 had slight differences between the two host temperatures. This implies that the differences in Flagellar assembly between the two host temperatures is largely due to the late stationary phase.

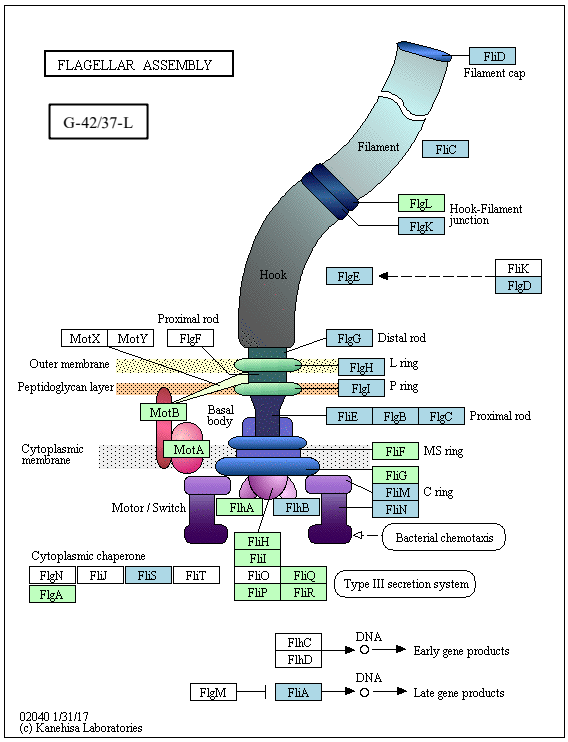


Figure 5.12: KEGG diagram of down-regulated genes that inactivate the Flagellar assembly pathway in 42 °C against 37 °C at late stationary phase. Blue boxes indicate the associated down-regulated genes. Green boxes indicate all the associated genes that are present in that organism. White boxes indicate the gene is present in the pathway but for the particular organism. Image taken from the KEGG database[3].

[3] `https://www.genome.jp/kegg/tool/map_pathway2.html`

In sub-cluster 1 it is clear that there are differences between the host temperatures dependent on growth phase. The general pathways: Metabolic pathways, Biosynthesis of amino acids, and Microbial metabolism in diverse environments were enriched from up-regulated genes in all comparisons. G-42/37-a ignores the effects of growth-phase so any enriched pathways observed in this comparison are due to differences between host temperatures. There were no enriched pathways from down-regulated genes for G-42/37-a as well as G-42/37-E and G-42/37-S and there is a cluster of enriched pathways from up-regulated genes for Nitrogen metabolism, Glyoxylate and dicarboxylate metabolism, Sulfur relay system, folate biosynthesis in which the latter is a part of the Molybdenum cofactor biosynthesis KEGG module. G-42/37-L which differs the most compared to the other comparisons had many enriched pathways from down-regulated genes including the aforementioned Flagellar assembly but also ABC transporters, Amino sugar and nucleotide sugar metabolism, and Thiamine metabolism. Many ABC transporter networks were enriched as seen in Figure 5.21, Section 5.4.7, although only gene network with the full set of genes Cj1276c and Cj1277c were present for ABC-2-type without transporter function.

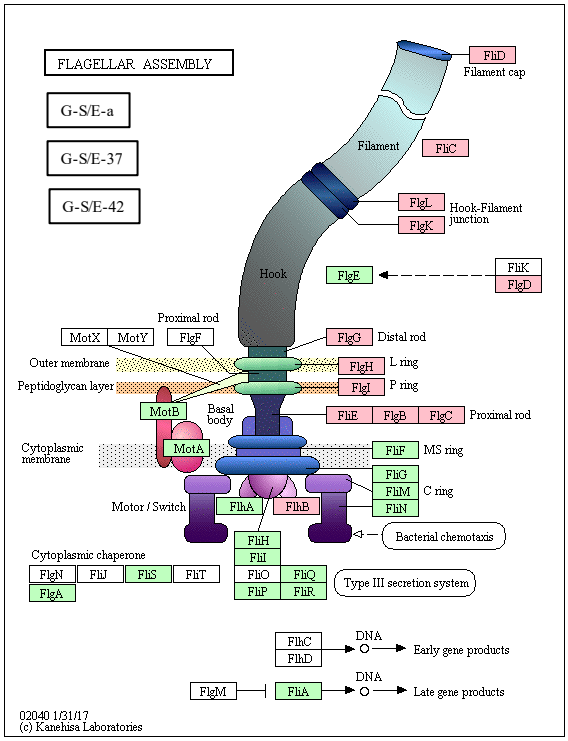


Figure 5.13: KEGG diagram of up-regulated genes that activate the Flagellar assembly pathway in all early stationary against exponential comparisons. Pink boxes indicate the associated up-regulated genes. Green boxes indicate all the associated genes that are present in that organism. White boxes indicate the gene is present in the pathway but not for that particular organism. Image taken from the KEGG database[3].

Sub-cluster 3 consists of all the late stationary against exponential phase comparisons and G-L/S-42 (42 °C late stationary against early stationary) (as was also seen in Figure 5.10) and is in a separate cluster with the other growth-phase dependent comparisons. This is due to many more pathways that were enriched mainly from down-regulated genes but also a few from up-regulated genes for all the comparisons. Many pathways enriched from down-regulated genes in all comparisons specific to the sub-cluster were under Carbohydrate metabolism. This is not surprising as cells transitioning into late stationary phase would be experiencing nutrient deprivation and growth would have reached saturation so there would be no need for the enzymes involved in metabolism to be expressed. The two-component system was also inactivated with the genes *frdABC*, *pstS*, *dcuB*, and *ccoPQO* common across the comparisons. However, *rpoN* which encodes for σ^{54} and also part of the Two-component system pathway was down-regulated in G-L/E-42. Some ABC transporters were inactivated though weakly and was not enriched in G-L/E-a which ignores host body temperature; the networks that were inactivated are shown in Figure 5.21, Section 5.4.7. As well as the Flagellar assembly, Ribosome and Valine, leucine and isoleucine biosynthesis pathways were also significantly enriched from up-regulated genes. ABC transporters were enriched from up-regulated genes albeit weakly in all comparisons except for G-L/S-42. There were only a few genes for particular gene sets that were present as seen in Figure 5.21, Section 5.4.7.

Sub-cluster 2 contains 37 °C late stationary against 37 °C early stationary phase (G-L/S-37) and all the late stationary conditions regardless of growth temperature against all early stationary phase conditions (G-L/S-a) and all transition from exponential to early stationary comparisons. But G-L/S-42 clustered separately in sub-cluster 3 suggesting there are many differences during this transition between the host temperatures. There were some pathways that were only enriched in the early stationary against exponential comparisons including ABC transporters and Valine, leucine and isoleucine biosynthesis from up-regulated genes, and Nitrogen metabolism and Terpenoid backbone biosynthesis from down-regulated genes. The ABC transporters full Iron(III) gene network *cfbpABC* was up-regulated across the early stationary against exponential comparisons implying that iron uptake is up-regulated during early stationary growth.

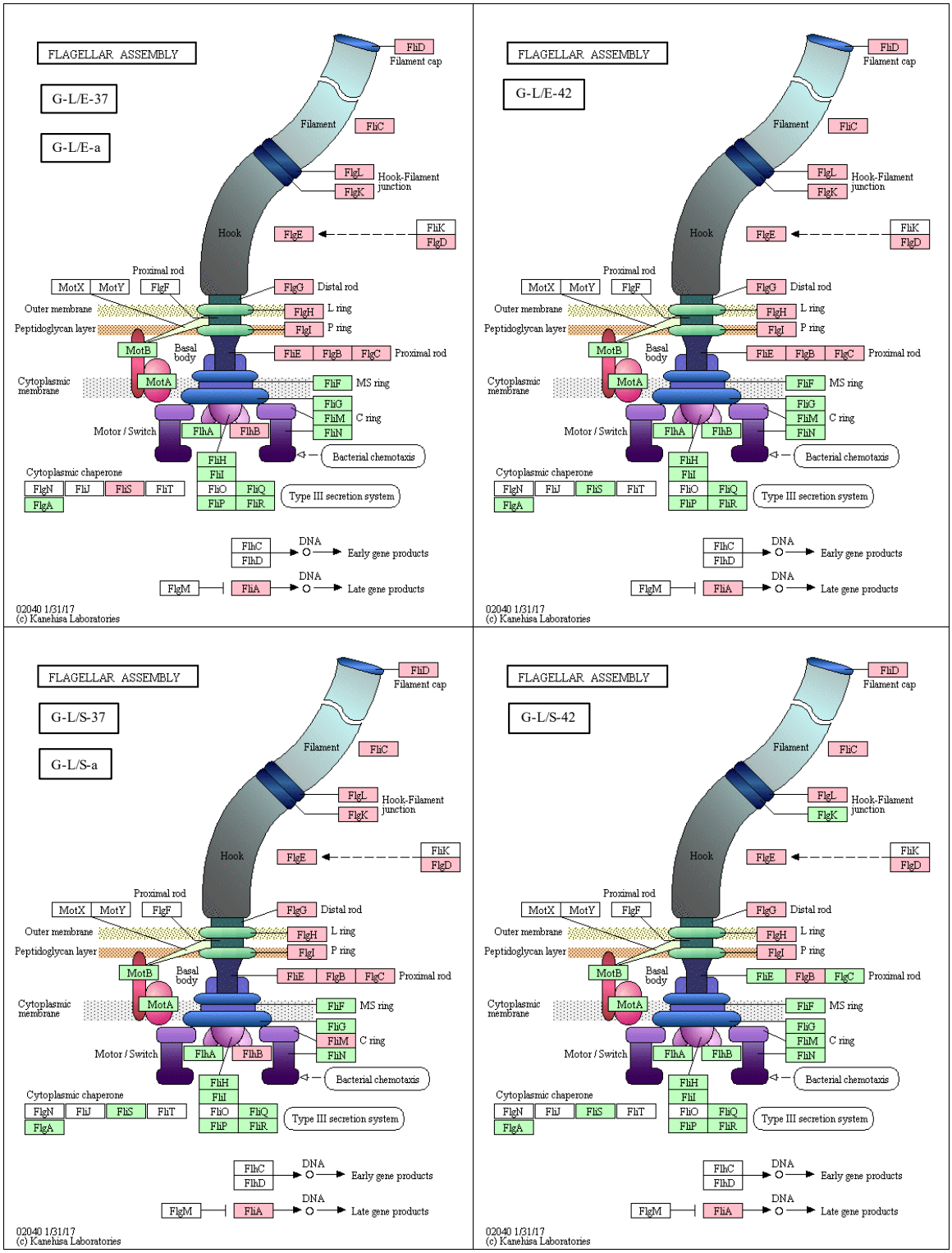


Figure 5.14: KEGG diagram of up-regulated genes that activate the Flagellar assembly pathway in late stationary against early stationary and exponential comparisons. Pink boxes indicate the associated up-regulated genes. Green boxes indicate all the associated genes that are present in that organism. White boxes indicate the gene is present in the pathway but not for that particular organism. Comparisons are shown in boxes. Image taken from the KEGG database[3].

[3] `https://www.genome.jp/kegg/tool/map_pathway2.html`

5.4.3 Effect of bile

C. jejuni is subjected to an environment with bile inside the host. Here bile refers to the addition of sodium deoxycholate, a component of bile as a supplement to MH2 media. NCTC 11168 was grown in bile and harvested at exponential and early stationary phase. Figure 5.15 is a heat map of the functional KEGG pathways enriched by differentially expressed genes from bile-related conditions. G-S/E-37 (37 °C early stationary against 37 °C exponential in MH2) was included as bile was compared against both of these conditions as seen in sub-clusters 1 and 2 of Figure 5.15. G-S/E-37 clusters more closely with B-b/37-E (sodium deoxycholate against 37 °C at exponential phase) in sub-cluster 1 as they down-regulate similar Metabolic pathways and strongly activates the Ribosome pathway. Sub-cluster 2 is sodium deoxycholate against 37 °C at early stationary (B-b/37-S) and at both growth phases combined (B-b/37-a). This sub-cluster is characterised by the strong activation of Metabolic pathways and strong inactivation of ABC transporters with moderate inactivation of Biosynthesis of amino acids. This sub-cluster clustered more closely with sub-cluster 3, which comprises of B-S/E-b (early stationary against exponential phase in the presence of bile) and B-S/E-a (early stationary against exponential independent of bile), which is defined by the strong enrichment of ABC transporters and moderate activation of the general pathway Biosynthesis of secondary metabolites by up-regulated genes.

There were major differences between B-b/37-E and B-b/37-S suggesting growth phase influences many changes in the presence of bile. At exponential phase, more pathways were down-regulated under Carbhydrate metabolism and Energy metabolism as highlighted by the orange box compared to early stationary phase, where more pathways were up-regulated under Energy metabolism and Amino acid metabolism as highlighted by the light blue box. There were also many pathways up-regulated in the transition from exponential to early stationary phase in the presence of bile only (B-S/E-b) including pathways under Amino acid metabolism and Replication and repair highlighted by the green box. Although ABC transporters were strongly inactivated in both B-b/37-E and B-b/37-S, different genes enriched for different networks. B-b/37-E had the entire Molydate, Tungstate, and Aspartate/Glutamate/Glutamine networks down-regulated whereas B-b/37-S only had a few genes that part of some gene networks as shown in Figure 5.21. On the other hand, B-S/E-b strongly activated the Iron(III), Tungstate, and Aspartate gene network as well as one gene in the Zinc and Molybdate networks.

The general pathway Carbon metabolism and specific pathways involved with Carbohydrate metabolism were only activated in B-S/E-b which is the inverse of B-b/37-E where many of the same pathways were inactivated. Some of the genes common across several of these pathways and part of the general Carbon metabolism pathway include *frdBC*, *sdhABC*, *pta*, *ackA*, *acnB*, *katA*, *acs*, *sdaA*, and *gltA*. There were some pathways which were common to both G-S/E-37 and B-b/37-E suggesting that the transition from exponential to early stationary phase in MH2 broth is more similar to the presence of bile at exponential phase than in early stationary phase. For example, both comparisons have the down-regulated genes *sdhABC* and *ppk* in Oxidative phosphorylation, although G-S/E-37 has many more genes enriching for that

pathway.

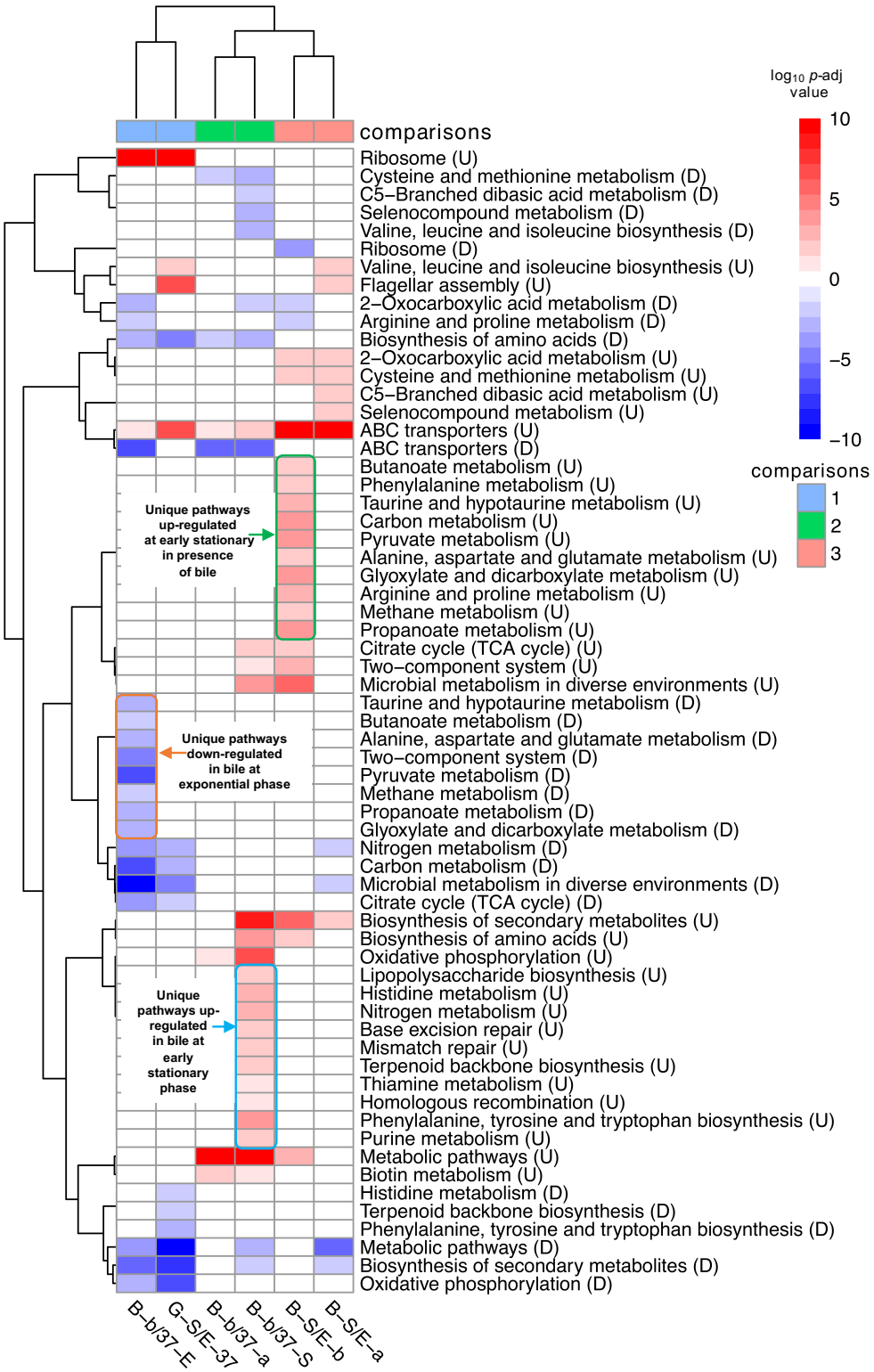


Figure 5.15: Heat map of functional enrichment KEGG pathway analysis of bile related comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes. Scale from red to blue indicates the intensity of the $\log_{10}$ *p*-adjusted value (inverse for up-regulated genes) and the comparisons legend indicates the sub-clusters.

5.4.4 Effect of iron limited and iron repleted conditions

Iron is an essential element required by all kingdoms of life for functioning cellular processes. But a paradox of iron usage is that too little or too much is both toxic to the cell prompting precise regulation to maintain iron homeostasis. The availability of iron can drastically change inside the host as an iron-limiting environment is a broad innate defence against pathogens, but nutrients released from consumption of food or from inflammation leads to ample availability of iron. *C. jejuni* was subjected to iron-limited and iron-repleted conditions in MEMα at both exponential and early stationary phase, and this was also compared with standard growth in MH2 broth, which is a rich media that contains beef extract and casein at both growth phases. Beef extract could be a possible source of iron. Figure 5.16 is a heat map of the different combinations of comparisons between iron limitation (IL), iron repletion (IR), and standard MH2 growth (37), and the significantly enriched KEGG pathways derived from up and down-regulated genes of these comparisons.

The comparisons were divided into four sub-clusters as seen by the annotated dendrogram. The general pathways: Metabolic pathways, Biosynthesis of secondary metabolites, and Biosynthesis of amino acids are general pathways that were enriched from up-regulated genes across all iron-related comparisons apart from the latter, which was not enriched in I-37/IR-a when ignoring the effect of growth phase between MH2 broth and MEMα supplemented with iron. There are two distinct clusters with sub-cluster 1 and 2 on the left, and sub-clusters 3 and 4 on the right. The difference between the two main clusters is due to the strong activation of pathways under Global and overview maps, Carbohydrate metabolism, and Energy metabolism in sub-clusters 1 and 2 as highlighted by the green box. ABC transporters were also down-regulated in all comparisons in these sub-clusters.

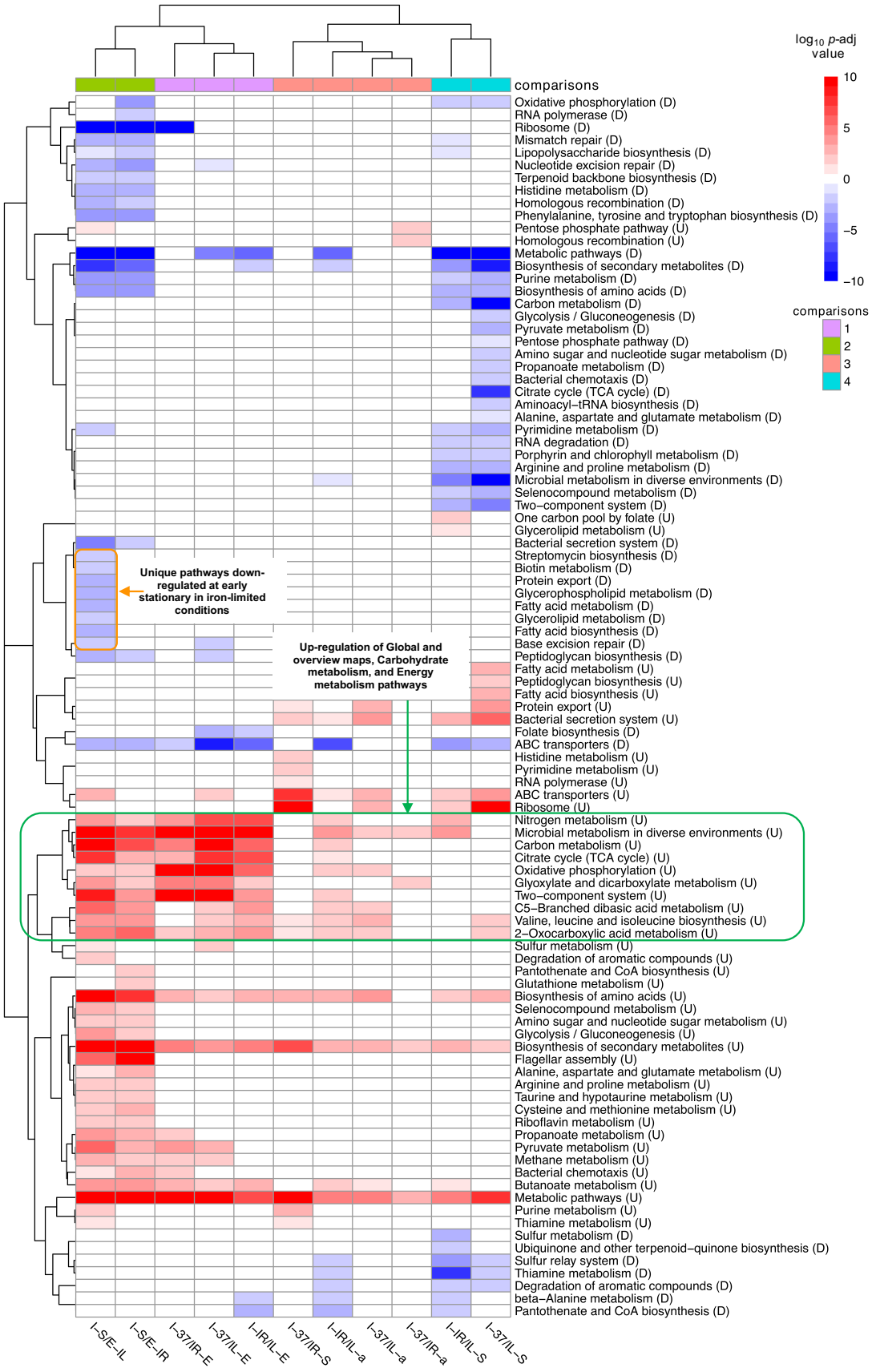


Figure 5.16: Heat map of functional enrichment KEGG pathway analysis of iron related comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes.

Sub-cluster 2 contains I-S/E-IL and I-S/E-IR, which is not surprising as both are early stationary against exponential phase in MEMα but with the latter supplemented with iron. There are slight differences though where in the absence of iron, many additional pathways such as Streptomycin biosynthesis, Biotin metabolism, Protein export, and the general Fatty acid metabolism pathway were down-regulated as highlighted by the orange box. Biotin metabolism is under Metabolism of co-factors and vitamins and it is possible that in the absence of iron many enzymes that require iron and other co-factors cannot function so these pathways would be down-regulated over time. In the presence of iron there were two pathways: RNA polymerase and Oxidative phosphorylation, which were enriched from down-regulated genes. Both comparisons also enriched for the Flagellar assembly from up-regulated genes due to the hierarchical nature of flagellar biogenesis. The genes involved in this pathway are shown in Figure 5.17. Several pathways related to Amino acid metabolism, Metabolism of other amino acids, and Carbohydrate metabolism were also up-regulated in this sub-cluster. The Two-component system pathway was also enriched from up-regulated genes and includes all three flagellin genes *flaABC*. A few pathways were enriched from down-regulated genes in sub-cluster 2 and in particular the Ribosome pathway which had a very strong *p*-adjusted value and Metabolic pathways. Some pathways were under Replication and repair and Amino acid metabolism. Lipopolysaccharide biosynthesis, Terpenoid backbone biosynthesis, and ABC transporters were also down-regulated.

Sub-cluster 1 is I-37/IR-E, clustered together with I-37/ILE and I-IR/IL-E. The latter two comparisons have iron limitation MEMα exponential phase as the control with rich or supplemented media as the "treatment". I-37/IR-E has compares MH2 against MEMα supplemented with iron. Many of the Energy metabolism pathways enriched from up-regulated genes in this sub-cluster was also enriched in sub-cluster 2. There were very few enriched pathways from down-regulated genes including ABC transporters across all comparisons in sub-cluster 1 but only Metabolic pathways and Folate biosynthesis were enriched in I-37/IL-E and I-IR/IL-E, which implies these pathways are up-regulated in iron-limited conditions. The Iron(III) and Phosphate ABC transporter networks were down in I-37/IR-E but some genes were missing from the set. But the entire Iron(III) gene set *cfbpABC* was down in both I-37/IL-E and I-IR/IL-E (Figure 5.21, Section 5.4.7). In addition, I-37/IL-E also had the entire Phosphate ABC transporter gene set *pstABCS* down. So during iron limitation the Iron(III) uptake system and the Phosphate uptake system is expressed, possibly due to the lack of phosphates in the minimal media MEMα.

Sub-cluster 3 has one individual pairwise comparison I-37/IR-S which clustered together with all the grouped comparisons I-IR-IL-a, I-37/IL-a, and I-37/IR-a that ignore the effect of growth phases. Only up-regulated genes significantly enriched for pathways for I-37/IR-S including ABC transporters and the Ribosome pathway, which has a very strong *p*-adjusted value, Histidine metabolism, Pyrimidine metabolism, RNA polymerase, Protein export, and Bacterial secretion system. The full set of genes for the ABC transporters Iron(III) uptake system *cfbpABC* and Branched-chain amino acid *livFGHJKM* were up-regulated in I-37/IR-S.

Sub-cluster 4 contains I-IR/IL-S and I-37/IL-S, which has iron limitation at early stationary as the control. Both comparisons enriched for similar pathways, although I-37/IL-S had many more Carbohydrate metabolism pathways that were inactivated again, likely due to the differences in minimal and rich media. There were also some pathways that were activated such as Fatty acid metabolism, Peptidoglycan biosynthesis, Fatty acid biosynthesis, and Protein export. I-IR/IL-S also had specific pathways that were enriched including Sulfur metabolism. Ubiquinone and other terpenoid-quinone biosynthesis, beta-Alanine metabolism, and Pantothenate and CoA biosynthesis from down-regulated genes. Both comparisons had many pathways enriched together including ABC transporters from both up and down-regulated genes. Both comparisons down-regulated the ABC transporter Phosphate uptake system but I-IR/IL-S up-regulated the entire Aspartate/Glutamate/Glutamine uptake system Cj0919-22c. I-37/IL-S did not have any gene networks that had the full list of genes present but still up-regulated various uptake systems with the missing genes possibly missing the cut-off. There were also genes that were down-regulated in both comparisons which enriched the same pathways including Cj0631c, *rho*, *groEL*, and *ppk* for RNA degradation and the flagellin genes *flaABC* in the Two-component system, although strangely the Flagellar assembly pathway was not enriched.

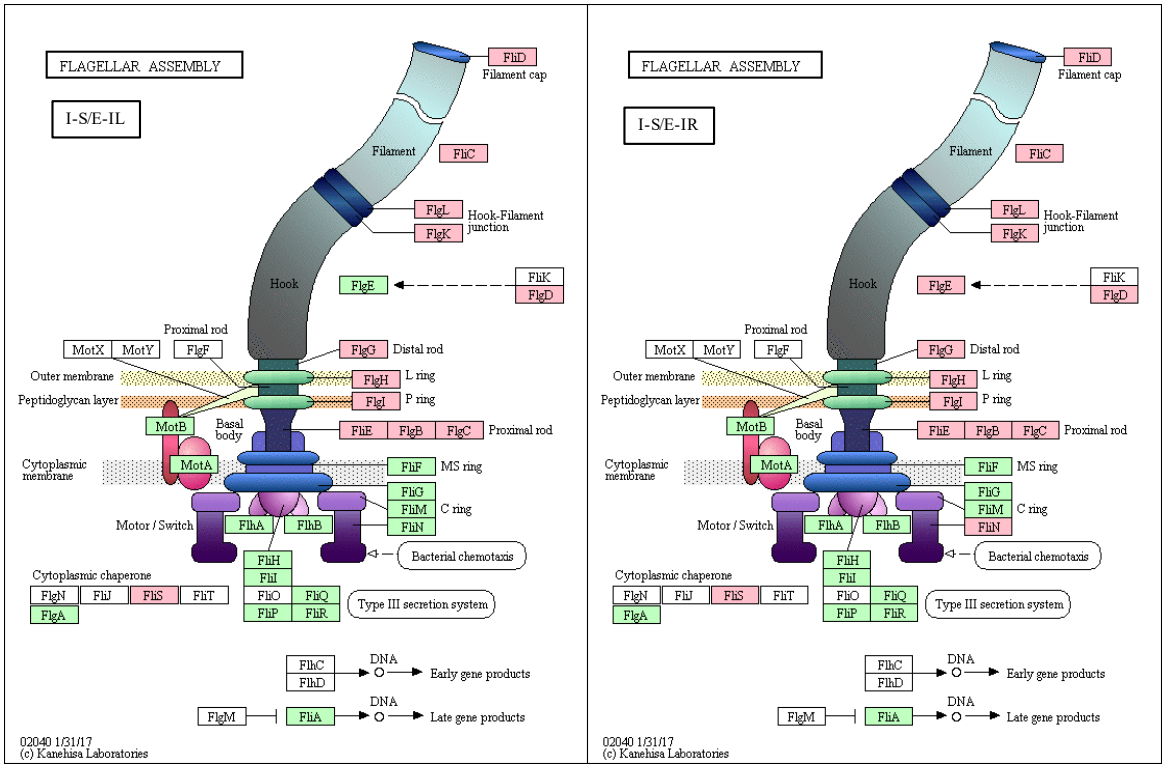


Figure 5.17: KEGG diagram of up-regulated genes that activate the Flagellar assembly pathway in iron-related comparisons. Pink boxes indicate the associated up-regulated genes. Green boxes indicate all the associated genes that are present in that organism. White boxes indicate the gene is present in the pathway but for the particular organism. Comparisons are denoted by the boxes. Image taken from the KEGG database[4].

[4] `https://www.genome.jp/kegg/tool/map_pathway2.html`

5.4.5 Effect of cold and chicken juice

C. jejuni is frequently found to contaminate retail poultry products such as chicken meat, which is primarily the source of transmission in developed countries (Hermans *et al.*, 2011b). These poultry products are stored at 4 °C to prolong longevity prior to consumption. NCTC 11168 was grown to exponential phase at 37 °C before incubation at 4 °C with: MH2 broth, diluted chicken juice (5 % v/v) in MH2 broth, and 100 % chicken juice for 24 hours. Therefore, 37 °C was used as a control for comparisons as well as cold and diluted chicken juice. Figure 5.18 is a heat map of all the pathways significantly enriched for cold-related comparisons.

There are two distinct main clusters in Figure 5.18 as seen on the annotated dendrogram. Cluster 1 contains all the 100 % chicken juice related comparisons: C-cj100/cj5, C-cj100/4, C-acj/E37, and C-cj100/E37, which strongly inactivated many pathways and only activated a few pathways in the latter two comparisons. As the RNA samples for 100 % chicken juice had very low coverage and low RIN values, this data is not trustworthy as the down-regulation of genes could be due to the low read counts. Hence, the data for 100 % chicken juice was largely ignored and not analysed in depth. Sub-cluster 2 contains the remaining cold-related conditions: C-cj5/4, C-cj5/E37, C-4/E37, and C-a4/E37. This cluster in particular activates many Amino acid metabolism pathways and Global and overview maps pathways across all the comparisons. The pathways were explored in more detail with the focus on conditions in sub-cluster 2.

C-cj5/E37 and C-4/E37 have a very similar pattern of significantly enriched KEGG pathways and are more closely clustered together. Even so, there are still different pathways enriched between the two comparisons and pathways enriched between C-cj5/4. C-a4/E37 (all the cold stresses combined) also has similar enrichment patterns to C-4/E37 although it also includes 100 % chicken juice. ABC transporters were enriched from up-regulated genes for all the comparisons in sub-cluster 2 apart from C-cj5/4. The entire Iron(III) network *cfbpABC* and *pstAC* from the Phosphate network enriched for the ABC transporters in all three comparisons, but for C-cj5/E37 the entire ABC transporter Zinc gene network and Cj1266/7c of the ABC-2-type components without transporting function was also up-regulated and enriched for this pathway.

The pathways common across all the comparisons in sub-cluster 2 were Global and overview maps pathways and many pathways under Amino acid metabolism enriched from up-regulated genes, the main cluster as highlighted by the light blue box, and C5-Branched dibasic acid metabolism. A few operons and genes were responsible for activating the pathways across the comparisons including *trpAB*, which significantly enriched for Glycine, serine and threonine metabolism, although *serC* was also present in the comparisons apart from C-cj5/4, and *pssA* was in C-cj5/E37 and C-4/E37, and *thrC* in C-cj5/E37. *leuABCD* significantly enriched for Biosynthesis of amino acids, Metabolic pathways, Biosynthesis of secondary metabolites, which all cluster together and solely enriched for Valine, leucine and isoleucine biosynthesis, and 2-oxocarboxylic acid metabolism, whereas *leuBCD* enriched for C5-Branched dibasic acid metabolism, though *leuA* was not present in C-cj5/4 and 2-oxocarboxylic acid was not enriched in C-cj5/E37. However,

C-cj5/E37 also up-regulated the entire Zinc network Cj0141-3c, and ABC-2-type without transporter function Cj1276c and Cj1277c. There were some enriched pathways from up-regulated genes exclusive to C-cj5/E37 and C-4/E37. Glycerophospholipid metabolism, Selenocompound metabolism, and Thiamine metabolism was enriched in C-cj5/E37 as highlighted by the purple box. C-4/E37 enriched for Carbon metabolism, Pyruvate metabolism, Alanine, aspartate and glutamate metabolism, and Microbial metabolism in diverse environments highlighted by the green box.

There were also pathways significantly inactivated by down-regulated genes in sub-cluster 2, although there were only two pathways Nitrogen metabolism enriched by *nrfAH* and Peptidoglycan biosynthesis enriched by *murDF* and *mraY*, which were down in C-cj5/4. The other comparisons in sub-cluster 2 all strongly enriched for Metabolic pathways from down-regulated genes, and weakly significant for pathways under Replication and repair, Lipopolysaccharide biosynthesis, and Ribosome pathway.

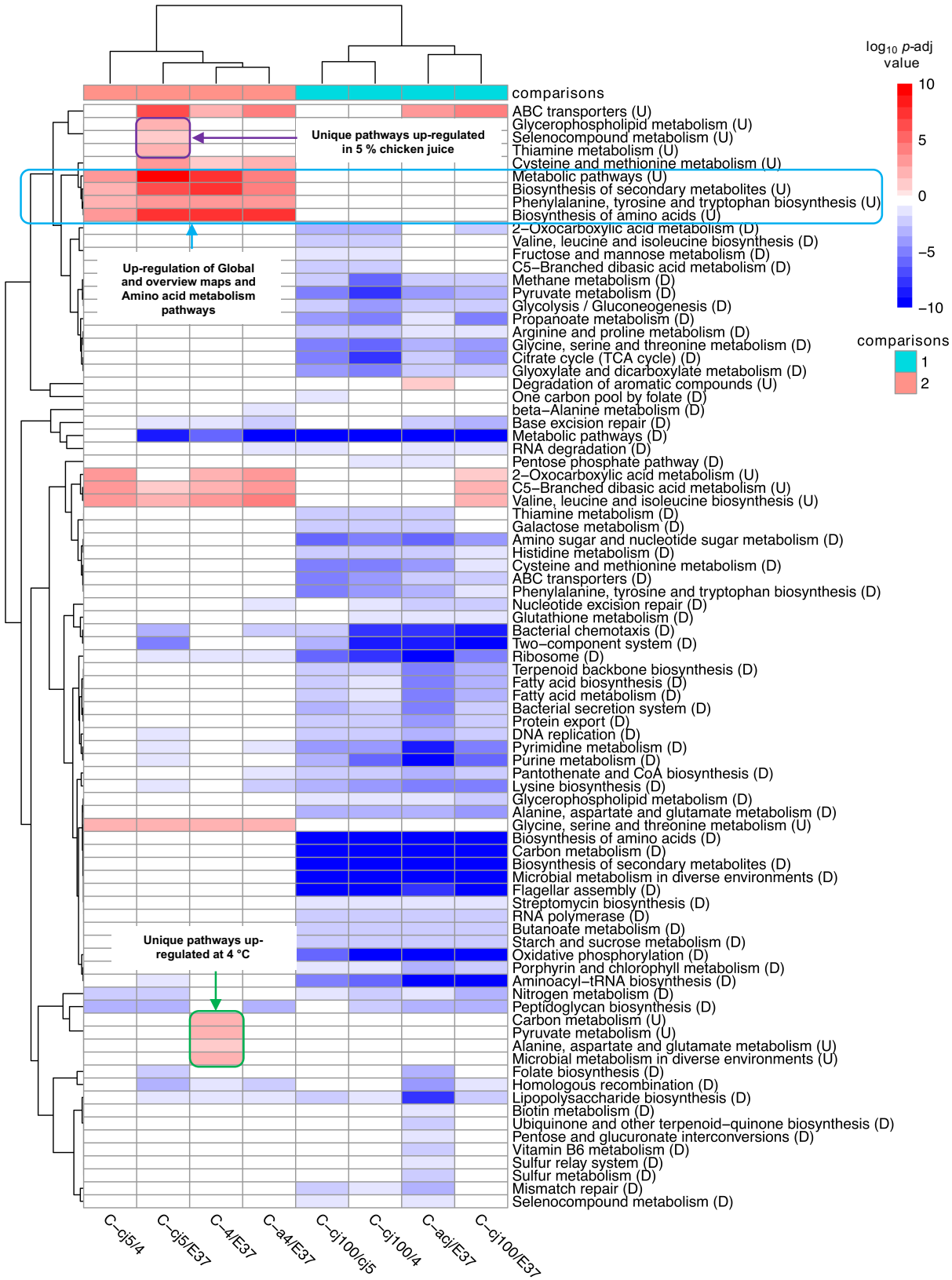


Figure 5.18: Heat map of chicken juice and cold related comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes.

5.4.6 Effect of various stresses

All stress shocks were compared against 37 °C exponential phase (E37) as this was the growth phase when the stresses were applied apart from starvation where 37 °C early stationary phase (S-starvation/S37) was used. Figure 5.19 is a heat map of all the stress comparisons enriching for KEGG pathways divided into four sub-clusters. Osmotic stress and starvation clustered separately from all the other stresses in their own sub-clusters, though starvation is closer to the other stresses. This was expected as both osmotic stress and starvation also clustered separately when comparing all the comparisons together in Figure 5.10. This is due to the many Metabolism pathways inactivated in osmotic stress including the branches of Global and overview maps, Carbohydrate metabolism, Energy metabolism, and amino acid metabolism highlighted by the orange box. Only a few enriched pathways were activated including RNA degradation and the Two-component system. Whereas starvation has many metabolic pathways that were enriched from up-regulated genes especially in Carbohydrate and Amino acid metabolism (highlighted by the green box) but no significantly enriched pathways from down-regulated genes. The other stresses were split into sub-clusters 1 and 2.

A distinctive feature of sub-clusters 1 and 2 is that very few pathways were enriched but ABC transporters were significantly activated by up-regulated genes as highlighted by the purple box. The common gene network that was activated across the stresses was *cfbpABC* of the Iron(III) network which uptakes ferric iron. Sub-cluster 2 which consists of anaerobic stress and GSNO clusters separately from sub-cluster 1 as it also strongly enriched for the Ribosome pathway from up-regulated genes. GSNO also inactivates a cluster of Metabolism pathways that were also strongly down-regulated in osmotic stress and inactivates the Molybdate ABC transporter gene network. It also weakly activated the Phenylalanine, tyrosine and tryptophan biosynthesis pathway with the operon *trpABDF* and *tyrA*. Anaerobic stress only had two pathways that were down-regulated: D-Glutamine and D-glutamate metabolism with *murI* and DNA replication from *rnhAB*. Anaerobic stress also activated RNA degradation by *dnaK*, *rho*, *groEL*, and Cj1710c; Valine, leucine and isoleucine biosynthesis by *leuABCD*; RNA polymerase by *rpoABC*; and the Flagellar assembly pathway with the associated genes shown in Figure 5.20.

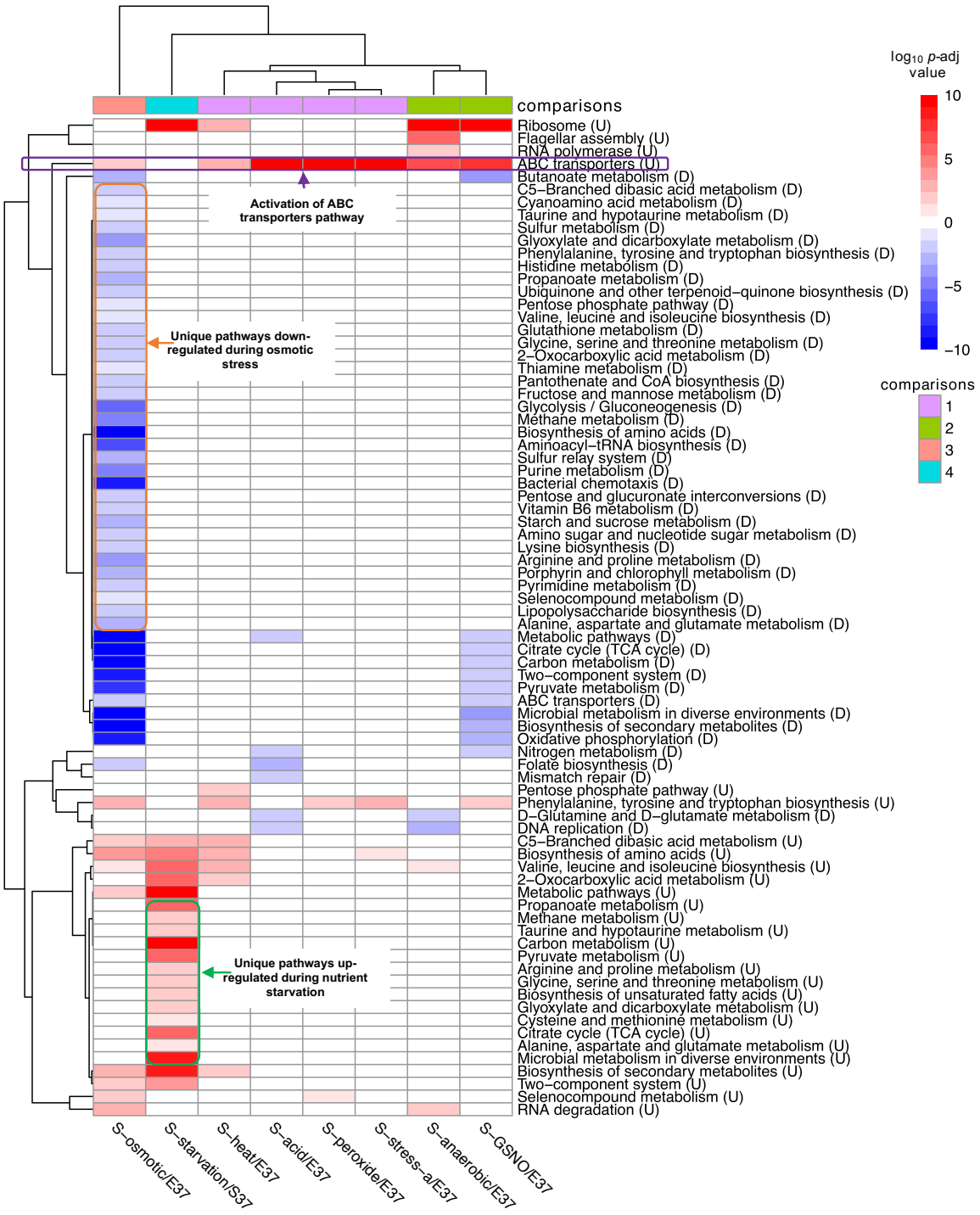


Figure 5.19: Heat map of stress shock related comparisons. Complete hierarchical clustering was used, with Euclidean distance for the comparisons, and Pearson correlation for the KEGG pathways. U = pathways enriched from up-regulated genes and D = pathways enriched from down-regulated genes.

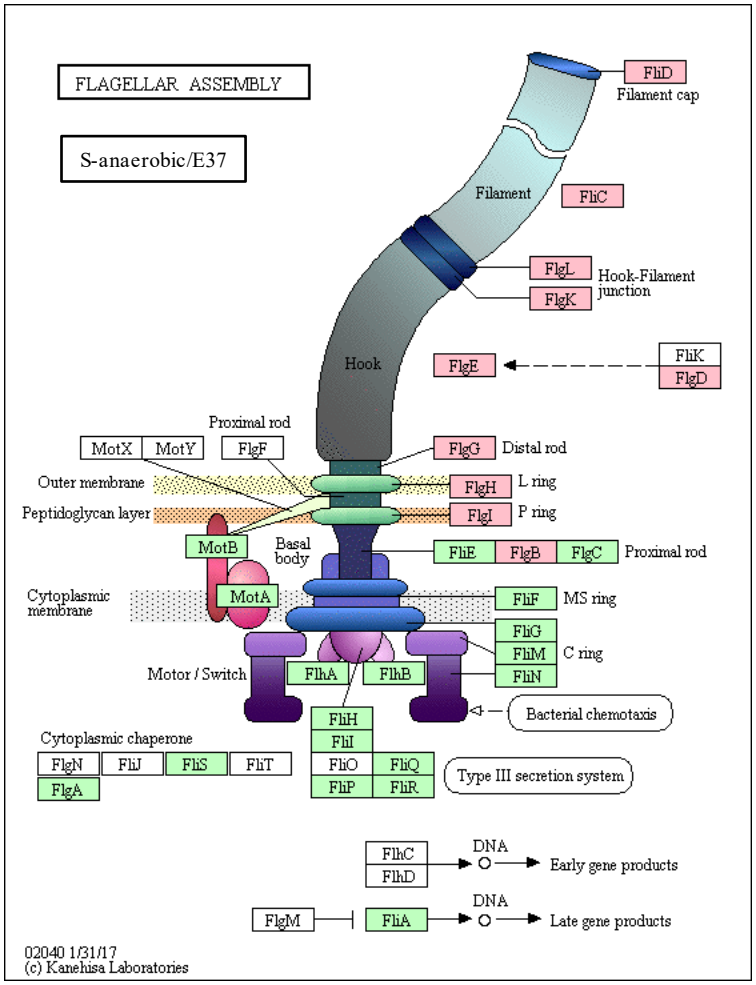


Figure 5.20: KEGG diagram of Flagellar assembly enriched in anaerobic stress compared against 37 °C exponential phase (S-anaerobic/E37). Pink boxes indicate the associated up-regulated genes. Green boxes indicate all the associated genes that are present in the organism. White boxes indicate the gene is generally present in the pathway but not for that particular organism.

Sub-cluster 1 contains heat shock, acid shock, peroxide stress, and all the stresses combined that were compared against 37 °C (i.e. excluding starvation). Overall stress enriches for the ABC transporter pathway Iron(III) gene network and *pstAC* from the phosphate network, Phenylalanine, tyrosine and tryptophan biosynthesis, and Biosynthesis of amino acids from *trpABDF* and the latter also from *leuD*. Both peroxide and heat shock had up-regulated enriched pathways only, whereas acid shock had down-regulated pathways apart from ABC transporters. Peroxide stress weakly enriched for Phenylalanine, tyrosine and tryptophan biosynthesis with *trpBF* and Selenocompound metabolism with *trxB*. Heat shock enriched for the Ribosome pathway with six genes encoding for the 50S subunit, and Biosynthesis of secondary metabolites from the operon *trpABDF* which also enriched Biosynthesis of amino acids, *leuBCD* which also enriched for Biosynthesis of amino acids; C5-Branched dibasic acid metabolism; Valine, leucine isoleucine biosynthesis; and 2-Oxocarboxylic acid metabolism, and finally *katA* catalase gene. Many genes are involved in the enriched pathways from down-regulated genes in acid shock. 20 genes are involved in Metabolic pathways though given the vast number of genes involved in this pathway the *p*-adjusted value was quite weak. Both mismatch repair and DNA replication were enriched by Cj0584, Cj0630c, Cj1669c,

and the latter also by *rnhB*. Nitrogen metabolism was enriched by *napB and nrfAH*, D-glutamine and D-glutamate metabolism by *murDI*, and Folate biosynthesis by *moaA*, *pabAB*, *folC* and *mobA* in which a few were also involved in Metabolic pathways.

Osmotic stress (sub-cluster 3) also enriches for pathways from up-regulated genes including ABC transporters *pstA* from the Phosphate network and Cj0313 and Cj0699 from the Lipopolysaccharide network. A number of up-regulated genes also enriched for Biosynthesis of amino acids including *trpABCD* which also enriched for the Phenylalanine, tyrosine and tryptophan biosynthesis, *leuCD* which also enriched for Valine, leucine and isoleucine biosynthesis and C5-Branched dibasic acid metabolism, *metE* which also enriched for Selenocompound metabolism, and finally *proB*. Specific up-regulated genes *dnak*, *rho*, and Cj1710c also activated RNA degradation, and *rpoN* also known as σ^{54}, *kdpB*, and Cj1361c activated the Two-component system. There were several pathways that were significantly inactivated by down-regulated genes especially Metabolic pathways which had 186 genes enriched for that pathway. The Molybdate, Tungstate, and Phospholipid ABC transporters were also inactivated.

Finally sub-cluster 4 is starvation compared with 37 °C exponential phase and had no enriched pathways from down-regulated genes and is the only stress that does not significantly enrich for ABC transporters. It strongly activates the Ribosome pathway, Metabolic pathways, Carbon metabolism, Microbial metabolism in diverse environments, and Biosynthesis of secondary metabolites and also activates many biosynthesis pathways related to amino acids. This is not surprising given the nutrient deprived environment and the reliance of amino acids as a carbon source for *C. jejuni*.

5.4.7 ABC transporters across 22 conditions

As ABC transporters pathways were consistently enriched in each group of comparisons, the gene networks were visually represented in Figure 5.21. For each gene network, the number of genes present was divided by the total number of known genes for that ABC transporter network in NCTC 11168, so a ratio of 1 indicates the entire gene network was enriched. Genes which were down-regulated were converted to negative numbers hence the scale of -1 to 1, where positive numbers are the presence of up-regulated genes (in red) and negative numbers are the presence of down-regulated genes (in green) and 0 implies there were no genes that were significantly enriched for the pathway or made the cutoff of > 1 $\log_2$ fold-change. There are two distinct clusters of comparisons divided into four sub-clusters, with the right cluster mainly containing down-regulated ABC transporter networks with a few up-regulated ABC transporters in the same comparisons and the left mainly containing up-regulated ABC transporters with some down-regulated transporters in the same comparison in sub-cluster 1. One pattern observed was that the Iron(III) and Other putative ABC transporter gene networks were either up-regulated or down-regulated together and there were no instances where Other putative was enriched by itself. There is only one gene present in Other putative ABC transporters Cj1587c, which is a multi-drug ABC transporter permease/ATP-binding protein.

Sub-cluster 1 contains comparisons that have ABC transporter networks enriched from both up and down-regulated genes with all comparisons down-regulating the Molybdate gene network. This sub-cluster includes all the comparisons comparing presence of bile to absence of bile in MH2 broth at both growth phases implying that Molybdate uptake is down-regulated in the presence of bile. GSNO against 37 °C also had the entire Molybdate gene network down-regulated and this comparison along with the bile comparisons also up-regulated the Phosphate and Phospholipid gene networks. The remaining comparisons are I-S/E-IL (transition from exponential to early stationary in MEMα in the absence of iron) and G-42/37-L (42 °C against 37 °C at late stationary phase), which had both up and down-regulated Molybdate uptake genes. All the comparisons also up-regulated the Phosphate uptake system except G-42/37-L.

Sub-cluster 3 only has comparisons that did not down-regulate any ABC transporters. All the comparisons also up-regulated the Iron(III) gene network except I-37/IL-a, which is MH2 broth against MEMα at both exponential and early stationary phase combined. All the cold-related comparisons are in this sub-cluster except C-cj5/4, which did not enrich for any ABC transporters. C-cj5/E37 (5 % chicken juice against 37 °C exponential phase) was the only comparison to up-regulate the ABC-2-type components without transporting function, which has the genes Cj1276c and Cj1277c. All stress shocks apart from osmotic stress and starvation are also in this sub-cluster. Peroxide stress, anaerobic stress, and B-S/E-a (transition from exponential to early stationary in the presence of bile independent of growth phase) cluster together with all the growth-related comparisons in sub-cluster 3, which are all the transition from exponential to early stationary phase at individual and combined host temperatures as they all up-regulated the full set of Iron(III) and Other putative ABC transporter networks. Phosphate uptake was also up-regulated for these comparisons though not the full gene set and not for B-S/E-a. G-L/E-a and I-37/IR-S also

clustered together and up-regulated Iron(III), Other putative, and partially Phosphate transporters but also up-regulated Macrolide exporters. Whereas B-S/E-b and I-37/IL-a clustered together as they both up-regulate the full Aspartate/Glutamate/Glutamine gene set.

Sub-cluster 4 contains all the individual late stationary against exponential and late stationary against early stationary comparisons at both host temperatures and osmotic stress, which is very similar to the overall clustering of all comparisons in sub-cluster 6 of Figure 5.10. The defining characteristics of sub-cluster 4 is the ABCB Subfamily gene network, which only has the one gene *msbA* that is down-regulated. Molybdate, D-Methionine, and Zinc gene networks were also down-regulated in all the comparisons, although not always the entire gene set. Sub-cluster 2 contain the remaining comparisons including those which did not enrich for any ABC transporters. G-L/S-a and I-S/E-IR clustered together and did not have any up-regulated ABC transporter networks. Both comparisons down-regulated Branched-chain amino acid, Molybdate, Tungstate, Phospholipid, and Zinc ABC transporter network. However, only I-S/E-IR down-regulated the full set of Branched-chain amino acid and Phospholipid networks. I-IR/IL-E and I-37/IR-E clustered together with the comparisons that had no enrichment and both down-regulated the Phosphate and Iron(III) gene network although I-IR/IL-E also down-regulated Other putative and the full set of Iron(III) network. MH2 broth compared to MEMα at both exponential and early stationary phase, and iron repletion against iron limitation at early stationary and both growth phases clustered together as they all down-regulated Phosphate, Iron(III), and Other putative ABC transporter networks but also up-regulated the Aspartate/Glutamate/Glutamine gene network.

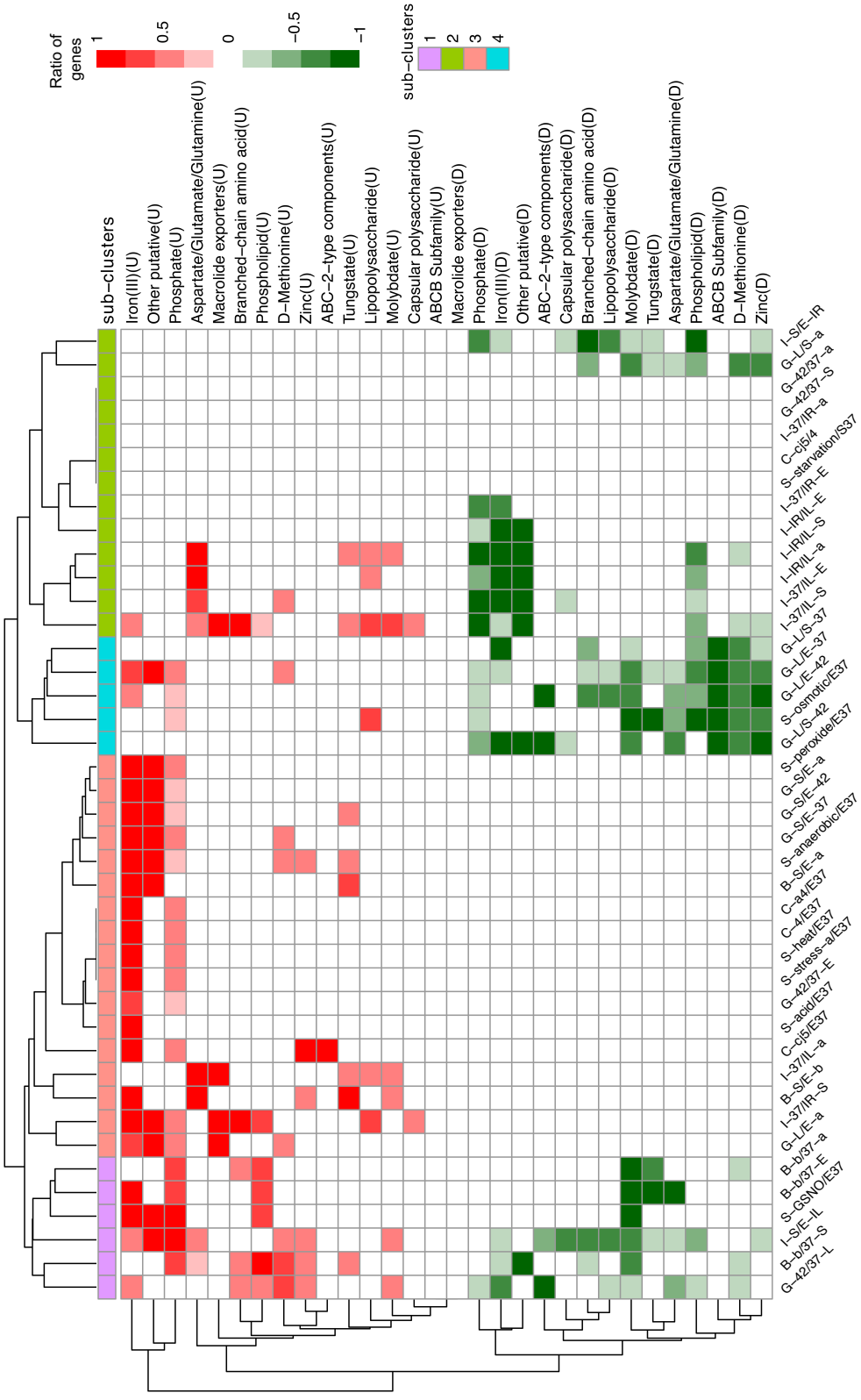


Figure 5.21: Heat map of the number of genes present in each gene network from enriched ABC transporters pathway in NCTC 11168. Complete hierarchical clustering with Euclidean distance was used. ABC-2-type components = ABC-2-type components without transporting function.

5.4.8 Variable fold-change

The most highly variable gene with the biggest lowest fold-change across all comparisons was Cj1384c in 37 °C early stationary against iron repletion early stationary control (I-37/IR-S) with 14.34 $\log_2$ fold-change and iron repletion early stationary against iron limitation early stationary control (I-IR/IL-S) with -15.01 $\log_2$ fold-change respectively. Cj1384c is an unknown hypothetical protein, however, it has been found to be highly differentially expressed in iron limited conditions or repressed in iron repleted conditions in Butcher and Stintzi (2013); Holmes *et al.* (2005); Palyada *et al.* (2004) and also expressed in sodium deoxycholate in Malik-Kale *et al.* (2008). It is divergently co-transcribed along with Cj1383c and the iron-binding protein FldA, antisense to *katA* in the antisense strand. Strangely, the ncRNA as_Cj1384c antisense to Cj1384c did not have any significant fold-changes across all comparisons.

5.5 The transcriptional response of osmotic stress

As osmotic stress exhibits unusual behaviour and clusters separately from all the other comparisons it was looked at individually in more detail. When put through the STRING database[5], the up-regulated genes (compared to 37 °C exponential as a control) enriched for the tryptophan biosynthesis process from the UniProt database. This corroborates with the functional enrichment analysis in Figure 5.19 where the pathways Biosynthesis of amino acids and Phenylalanine, tyrosine and tryptophan biosynthesis had the highest *p*-adjusted values. Genes encoding for tryptophan biosynthesis *trpABDF* were the genes that enriched for the pathway. This also correlates with a previous study on hyperosmotic stress by Cameron *et al.* (2012) carried out in strain 81-176, which also found many amino acid biosynthesis genes including tryptophan biosynthesis up 3-fold (Cameron *et al.*, 2012).

Surprisingly, many ncRNAs especially antisense RNAs were significantly up-regulated in osmotic stress. Furthermore, osmotic stress had the most identified ncRNAs from Chapter 4. A total of 452 gene features were up-regulated with 337 ncRNAs with *p*-adjusted value ≤ 0.05 and $\log_2$ fold-change ≥ 1. Whereas 750 gene features including 116 ncRNAs were down-regulated. The top ten most highly up-regulated genes are shown in Table 5.7 with the top five all ncRNAs. The top hit is an intergenic ncRNA int_Cj0747 upstream of Cj0747, an unknown gene identified from Chapter 4 using toRNAdo. The genes Cj0747 and Cj0748 which are further downstream of int_Cj0747 are also on the list suggesting that int_Cj0747 is highly likely a 5'UTR. Though the functions of Cj0747 and Cj0748 are unknown, both are only specific to *C. jejuni* and no homologues have been found outside of this species (Metris *et al.*, 2011).

All the other ncRNAs are antisense and were identified from Dugar *et al.* (2013)'s study. as_Cj0053c is antisense to Cj0053c or *trmU* which is a probable tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase. *trmA* is another probable tRNA-methyltransferase, which is down-regulated in osmotic stress (Table 5.7). However, *trmU* only has a slight significant $\log_2$ fold-change of -0.6 so it is not clear whether tRNA-methyltransferase in general is down-regulated. Many of the down-regulated genes also include probable integral or periplasmic proteins (Table 5.7).

[5] https://string-db.org/cgi/input.pl?sessionId=LPYUoO1eV1aq&input_page_show_search=on

Locus tag	Gene	Gene product/description	Log_2 fold-change	Adjusted *p*-value
Up-regulated				
int_Cj0747	-	possible 5' UTR of Cj0747	8.09	4.47E-33
as_Cj0053c	-	-	7.93	4.70E-05
as_Cj1395_1	-	-	7.75	0.0002671
as_Cj0992c	-	-	7.56	0.00350834
Cj0748	-	unknown	7.52	1.03E-54
as_Cj0287c	-	-	7.50	0.00011272
as_Cj0639	-	-	7.46	0.00068726
Cj0752;part=2/3;pseudo=true	-	probable IS element transposase pseudogene	7.38	7.24E-54
Cj0747	-	unknown	7.36	2.85E-32
as_Cj0638c	-	-	7.30	0.00086281
Down-regulated				
Cj0659c	-	possible periplasmic protein	-5.34	1.99E-14
Cj0831c	*trmA*	probable tRNA(uracil-5-)-methyltransferase	-5.23	4.54E-75
Cj1078	-	possible periplasmic protein	-4.88	1.55E-13
Cj0300c	*modC*	probable molybdenum transport ATP-binding protein	-4.79	1.31E-51
Cj0832c	-	probable integral membrane protein	-4.48	7.71E-63
Cj0073c	-	unknown	-4.45	6.85E-38
Cj0481	*dapA*	probable lyase	-4.42	3.17E-31
Cj0482	*uxaA*	possible altronate hydrolase N-terminus	-4.29	1.48E-09
Cj0911	-	probable periplasmic protein	-4.22	8.86E-17
Cj0833c	-	probable oxidoreductase	-4.19	1.03E-42

Table 5.7: List of ten most highly differentially regulated genes in osmotic stress compared to 37 °C exponential phase.

5.6 Normalisation with TPM

Transcripts per million (TPM) were calculated for all read counts obtained using coverageBed for genes, pseudogenes, and ncRNAs to normalise for gene length and depth (Wagner *et al.*, 2012). CjNC19, an intergenic ncRNA found in Porcelli *et al.* (2013) did not have any gene counts at all across all three replicates of 22 conditions i.e. 66 samples. Decile values for each sample were calculated from TPM values to categorise highly and lowly expressed genes as was done in Aprianto *et al.* (2018). As there was low coverage for 100 % chicken juice condition, the samples were not included for determining highly and lowly expressed genes as it would skew the data. 61 gene features including 11 ncRNAs were found to be highly expressed i.e. values ranked in the tenth decile partition in all samples (Table 5.8). Many of these genes encode for rRNA or are involved in the structure of the membrane but only nine were essential when compared to the 166 essential genes in NCTC 11168 determined by Mandal *et al.* (2017). The origin of the 11 highly expressed ncRNAs and description was listed in Table 5.9. Many of the ncRNAs overlapped genes and were all identified by Dugar *et al.* (2013). There were no genes that were lowly expressed in the bottom decile in all conditions (not including 100 % chicken juice). 1508 gene features were found to be conditionally expressed where in at least one condition, the TPM values were in the bottom decile i.e. first partition. The number of genes and ncRNAs that were highly or conditionally expressed are shown in Figure 5.22.

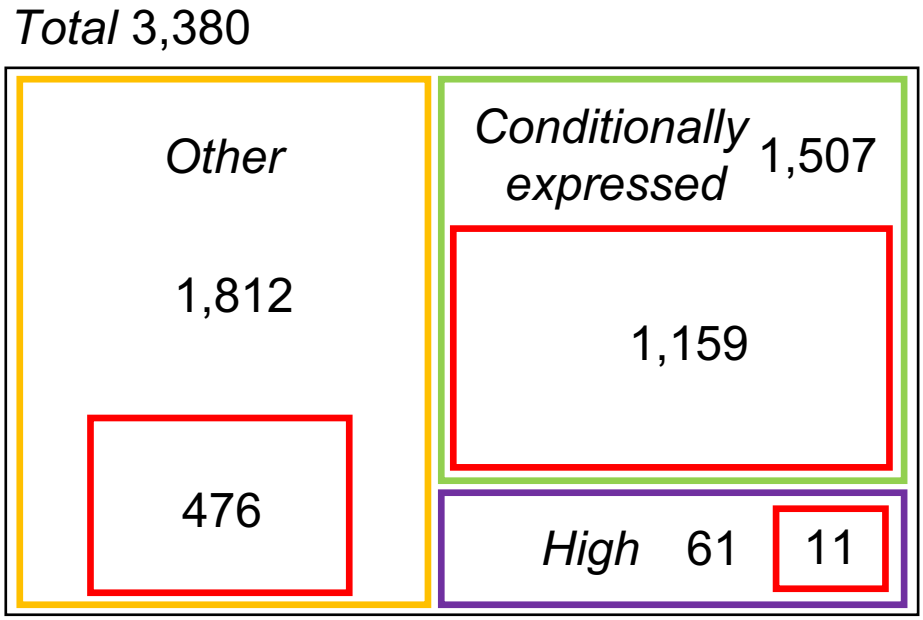


Figure 5.22: Categorisation of gene features by TPM values in deciles. Red boxes indicate the number of ncRNAs in each group. 'High' indicates all gene features which were highly expressed in all conditions present in the top decile, 'Conditionally expressed' are all gene features that were in the bottom decile in at least one condition, and 'Other' are the remaining gene features which did not fit these criteria.

Locus tag	Gene name	Product description	Essential
Cj0094	*rplU*	50S ribosomal protein L21	Yes
Cj0095	*rpmA*	50S ribosomal protein L27	No
Cj0113	*pal*	peptidoglycan associated lipoprotein	Yes
Cj0147c	*trxA*	thioredoxin	No
Cj0236c	-	integral membrane protein	No
Cj0330c	*rpmF*	50S ribosomal protein L32	Yes
Cj0334	*ahpC*	alkyl hydroperoxide reductase	No
Cj0370	*rpsU*	30S ribosomal protein S21	Yes
Cj0391c	-	hypothetical protein Cj0391c	No
Cj0408	*frdC*	fumarate reductase cytochrome B subunit	No
Cj0410	*frdB*	fumarate reductase iron-sulfur subunit	No
Cj0420	-	periplasmic protein	No
Cj0428	-	hypothetical protein Cj0428	No
tRNA-Thr_gene444	-	-	No
Cj0470	*tuf*	elongation factor Tu	No
Cj0526c	*fliE*	flagellar hook-basal body protein FliE	Yes
Cj0527c	*flgC*	flagellar basal body rod protein FlgC	No
as_Cj0551	-	-	No
RnpB	-	-	No
rnpB_gene526	-	-	No
as_Cj0595c	-	-	No
Cj0628;part=1/2	-	lipoprotein	No
Cj0699c	*glnA*	glutamine synthetase	No
Cj0715	-	5-hydroxyisourate hydrolase	No
Cj0716	-	phospho-2-dehydro-3-deoxyheptonate aldolase	No
Cj0720c	*flaC*	flagellin C	No
Cj0772c	-	NLPA family lipoprotein	No
Cj0779	*tpx*	thiol peroxidase	No
Cj0913c	*hupB*	DNA-binding protein HU	No
Cj0936	*atpE*	ATP synthase subunit C	No
Cj0998c	-	periplasmic protein	No
Cj1026c	-	lipoprotein	No
Cj1070	*rpsF*	30S ribosomal protein S6	Yes
Cj1071	*ssb*	single-stranded DNA-binding protein	Yes
Cj1072	*rpsR*	30S ribosomal protein S18	No
Cj1110c	-	MCP-type signal transduction protein	No
Cj1118c	*cheY*	chemotaxis protein CheY	No

Table 5.8 continued from previous page

Locus tag	Gene name	Product description	Essential
Cj1153	-	cytochrome C	No
Cj1176c	*tatA*	Sec-independent protein translocase	Yes
as_Cj1219c_1	-	-	No
as_Cj1219c_2	-	-	No
Cj1220	*groES*	co-chaperonin GroES	Yes
Cj1221	*groEL*	chaperone GroEL	No
CJnc120	-	-	No
Cj1242	-	hypothetical protein Cj1242	No
6S_RNA_(CJnc130)	-	-	No
as_Cj1257c	-	-	No
Cj1259	*porA*	major outer membrane protein	No
Cj1338c	*flaB*	flagellin B	No
Cj1339c	*flaA*	flagellin A	No
as_Cj1362	-	-	No
tmRNA	10Sa RNA (tmRNA)	putative proteolysis tag	No
Cj1360c;partial=true	-	putative proteolysis tag	No
as_Cj1361	-	-	No
Cj1419c	-	methyltransferase	No
Cj1420c	-	methyltransferase	No
Cj1464	*flgM*	flagellar biosynthesis protein FlgM	No
Cj1476c	-	pyruvate-flavodoxin oxidoreductase	No
Cj1490c	*ccoN*	cbb3-type cytochrome c oxidase subunit I	No
Cj1534c	-	bacterioferritin	No
Cj1625c	*sdaC*	amino acid transporter	No

Table 5.8: List of highly expressed genes in all conditions. The highly expressed genes were compared with a list of essential genes by Mandal *et al.* (2017).

ncRNA	TSS			Notes
	Dugar *et al*, 2013	Porcelli *et al*, 2013	Cappable-seq	
as_Cj0551				on top of tmRNA
rpnB_gene526				promoter
as_Cj0595c				on top of peb4-cbf2
as_Cj1219c_1				runs into *groES* and *groEL*
as_Cj1219c_2				runs into *groES* and *groEL*
6S_RNA_(CJnc130)				house-keeping RNA
as_Cj1257c				on top of Cj1258 and *porA*
as_Cj1362				on top of Cj1360c and Cj1361c
as_Cj1361				on top of Cj1361c
tmRNA				also known as 10s RNA
CJnc120				on top of *groEL*

Table 5.9: List of highly expressed ncRNAs. Red - Present. Green - Absent.

5.6.1 The behaviour of stringent response related genes across 22 conditions

The stringent response is heavily involved in regulating the bacterial stress response in a number of nutrient-limited conditions (Irving and Corrigan, 2018). Since many of the host-relevant and transmission conditions used in this project involve stress, the expression of stringent response related genes were investigated for any patterns and trends. The full list of genes selected are shown in Table 5.10. (p)ppGpp is an alarmone signalling molecule responsible for triggering the stringent response and high concentrations of this molecule lead to repression of many biosynthesis genes; this is regulated by *spoT* (Gaca *et al.*, 2015). The interactions of *spoT* with other genes and proteins (excluding neighbourhood genes as strandedness is not accounted for) were retrieved from the STRING database website[6]. All genes with a high confidence interaction with *spoT* as seen in Figure 5.23 were included in the list. The homologues of polyphosphate genes *ppk1* and *ppk2* and associated genes were also included as they have been implicated as global regulators of bacterial stress (Kumar *et al.*, 2016). Other genes such as *neuA1,cgtA* and *acpP* were found in the literature to associate with *spoT* (Irving and Corrigan, 2018). The rationale for manually including adjacent genes and antisense ncRNAs in the vicinity was that genes within an operon under a single promoter would be expected to have similar expression levels and antisense ncRNAs could possibly target the genes on the sense strand, therefore it would be informative to also include these gene features. These genes were manually selected as the neighbourhood option in STRING does not account for strandedness and we also have identified additional predicted promoters from Chapter 4.

[6] `https://string-db.org/`

Stringent response gene/locus tag	Operon	ncRNAs	Description	Source
spoT	Cj1268c	-	bifunctional tRNA (mnm(5)s(2)U34)-methyltransferase/ FAD-dependent cmnm(5)s(2)U34 oxidoreductase	
	amiA (Cj1269c)	as_Cj1269c	N-acetylmuramoyl-L-alanine amidase	
	Cj1270c	as_Cj1270c	2-nitropropane dioxygenase	
	tyrS (Cj1271c)	as_Cj1271c	tyrosine–tRNA ligase	
	spoT (Cj1272c)	as_Cj1272c_1 as_Cj1272c_2	guanosine-3' ,5'-bis(diphosphate) 3'-pyrophosphohydrolase	(Gaynor *et al.*, 2005)
	rpoZ (Cj1273c)	-	DNA-directed RNA polymerase subunit omega	STRING database
	pyrH (Cj1274c)	-	uridylate kinase	
ppk	**ppk (Cj1359)**	as_Cj1359_2 as_Cj1359_3	polyphosphate kinase	(Kumar *et al.*, 2016)
Cj0604 (ppk2)	**Cj0604 (ppk2)**	as_Cj0604_2	polyphosphate kinase	(Kumar *et al.*, 2016)
	Cj0605	-	amidohydrolase	
lon	**lon (Cj1073c)**	-	ATP-dependent protease La	(Cohn *et al.*, 2007)
	Cj1074c	as_Cj1074 as_Cj1075	lipoprotein	
	purN (Cj0187c)	as_Cj0186c as_Cj0187c_1 as_Cj0187c_2	phosphoribosylglycinamide formyltransferase	
	Cj0188c	-	kinase	
	Cj0189c	-	unknown (hypothetical protein)	

Table 5.10 continued from previous page

Stringent response gene/locus tag	Operon	ncRNAs	Description	Source
clpP	Cj0190c	as_Cj0190c_1 as_Cj0190c_2	unknown (hypothetical protein)	
	def (Cj0191c)	-	peptide deformylase	
	clpP (Cj0192c)	as_Cj0192c_1 as_Cj0192c_2	ATP-dependent Clp protease proteolytic subunit	(Cohn *et al.*, 2007)
	Cj0193c	-	trigger factor	
neuA1,cgtA	cstIII (Cj1140)	as_Cj1140_1 as_Cj1140_2	alpha-2,3 sialyltransferase	
	neuB (Cj1141)	-	sialic acid synthase	
	neuC1 (Cj1142)	as_Cj1142_1 as_Cj1142_2	UDP-N-acetylglucosamine 2-epimerase	
	neuA1, cgtA (Cj1143)	as_Cj1143_1 as_Cj1143_2	bifunctional beta-1,4-N-acetylgalactosaminyltransferase/CMP-Neu5Ac synthase	(Irving and Corrigan, 2018)
rpoB	**rpoB (Cj0478)**	-	DNA-directed RNA polymerase subunit beta	STRING database
	rpoC (Cj0479)	as_Cj0479_1 as_Cj0479_2	DNA-directed RNA polymerase subunit beta'	
Cj0353c (ppx/gppA)	**Cj0353c (ppx/gppA)**	as_Cj0353c	phosphatase	STRING database, (Malde *et al.*, 2014)
	fdxB (Cj0354c)	-	ferredoxin	

Table 5.10 continued from previous page

Stringent response gene/locus tag	Operon	ncRNAs	Description	Source
acpP	**acpP (Cj0441)**	as_Cj0441c as_Cj0440c_1 as_Cj0440c_2	acyl carrier protein	(Angelini *et al.*, 2012), (Irving and Corrigan, 2018)
	fabF (Cj0442)	-	3-oxoacyl-ACP synthase	(Irving and Corrigan, 2018)
	accA (Cj0443)	as_Cj0444_1 as_Cj0444_2	acetyl-CoA carboxylase carboxyltransferase subunit alpha	
clpX	fabZ (Cj0273)	-	3-hydroxyacyl-ACP dehydratase	(Cohn *et al.*, 2007)
	lpxA (Cj0274)	as_Cj0274_1 as_Cj0274_2	acyl-ACP–UDP-N-acetylglucosamine O-acyltransferase	
	clpX (Cj0275)	-	ATP-dependent protease ATP-binding subunit ClpX	
	mreB (Cj0276)	-	rod shape-determining protein MreB	
	mreC (Cj0277)	as_Cj0279	rod shape-determining protein MreC	
pnp	**pnp (Cj1253)**	-	polynucleotide phosphorylase	STRING database
	Cj1254	-	unknown (hypothetical protein)	
gmk	argS (Cj1175c)	as_Cj1175c	arginine–tRNA ligase	STRING database
	tatA (Cj1176c)	-	Sec-independent protein translocase	
	gmk (Cj1177c)	-	guanylate kinase	
	Cj1178c	as_Cj1178c_1	highly acidic protein	

Table 5.10 continued from previous page

Stringent response gene/locus tag	Operon	ncRNAs	Description	Source
nrdA	purB (Cj0023)	as_Cj0022c as_Cj0023c	adenylosuccinate lyase	
	nrdA (Cj0024)	as_Cj0024	ribonucleotide-diphosphate reductase subunit alpha	STRING database
nrdB	Cj0230c	as_Cj0230c	nicotinate phosphoribosyltransferase	
	nrdB (Cj0231c)	-	ribonucleotide-diphosphate reductase subunit beta	STRING database
	Cj0232c	as_Cj0232c	integral membrane protein	
	pyrE (Cj0233c)	-	orotate phosphoribosyltransferase	
	frr (Cj0234c)	-	ribosome recycling factor	
	secG (Cj0235c)	-	protein translocase subunit SecG	
rpoA	rpmJ (Cj1591)	-	50S ribosomal protein L36	
	rpsM (Cj1592)	-	30S ribosomal protein S13	
	rpsK (Cj1593)	-	30S ribosomal protein S11	
	rpsD (Cj1594)	-	30S ribosomal protein S4	
	rpoA (Cj1595)	-	DNA-directed RNA polymerase subunit alpha	STRING database
	rplQ (Cj1596)	as_Cj1597_1	50S ribosomal protein L17	
Cj1237c	**Cj1237c**	-	phosphatase	STRING database, (Malde *et al.*, 2014)

Table 5.10 continued from previous page

Stringent response gene/locus tag	Operon	ncRNAs	Description	Source
pyk	**pyk (Cj0392c)**	as_Cj0392c_1 as_Cj0392c_2 as_Cj0392c_3	pyruvate kinase	STRING database

Table 5.10: List of stringent response genes and their function and source.

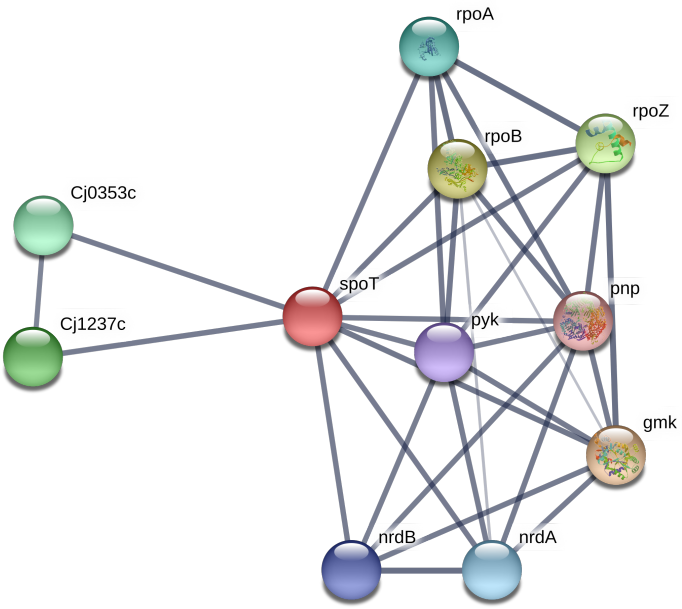


Figure 5.23: Network of known and predicted interactions with *spoT*. Generated from STRING database. The thickness of network edges reflects the confidence score.

A heat map of $\log_2$ TPM values against stringent response related genes is shown in Figure 5.24 with the short-hand key in Table 5.5. Most of the replicates from each condition clustered together apart from one replicate from each of the following conditions: GSNO, cold stress, iron repletion exponential phase (IRE), and sodium deoxycholate early stationary replicates (bS). The conditions were split into seven colour-coded clusters denoted by the annotated dendrogram at the top. The iron limitation early stationary (ILS) replicates of cluster 4 clustered separately from all the other conditions, as the expression of a block of genes from sections I and III-V that were highly expressed in all other conditions were not as highly expressed in ILS. It was also interesting to see the stringent response genes behaving in a similar way in nutrient deprived early and late stationary conditions as they all cluster together in clusters 5 and 7 and are distinct from the other clusters due to the low TPM expression of genes in section V as highlighted by the red box. Cluster 5 contains the standard growth conditions in each host temperature at early stationary phase 37 °C (S37) and 42 °C (S42) carried out in MH2 broth. Whereas cluster 7 has the late stationary phase conditions at 37 °C (L37), 42 °C (L42), starvation, and iron repletion early stationary phase (IRS). The only other early stationary phase condition is in sodium deoxycholate (bS) where the replicates clustered separately from each other in cluster 1.

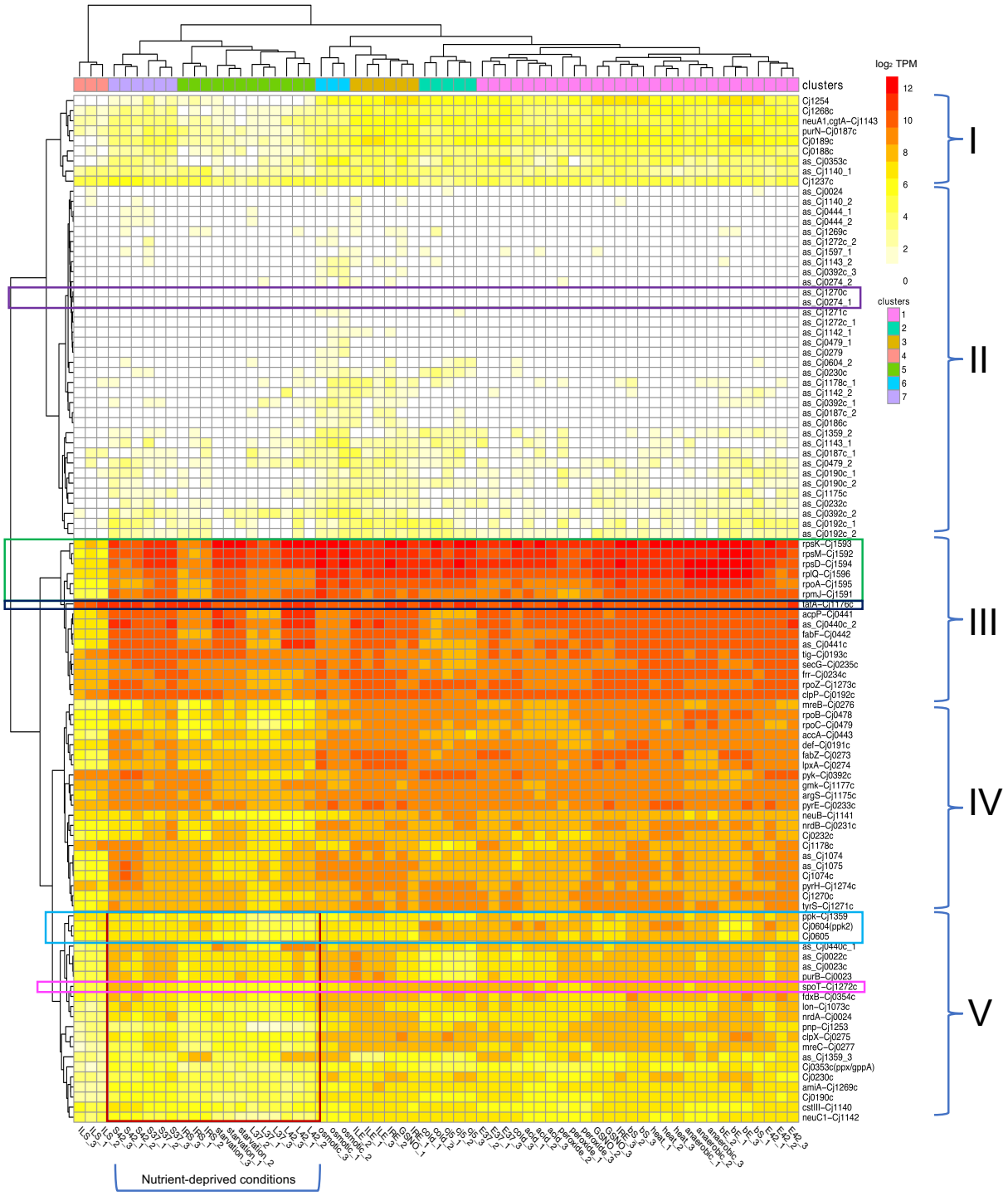


Figure 5.24: Heat map of absolute expression of stringent response genes. Absolute expression values are $\log_2$ normalised TPM values. Euclidean distance with hierarchical clustering was used. Important genes and sections are annotated in coloured boxes and explained in detail in the main text.

Cluster 6 contains just the osmotic stress replicates, which clustered together with cluster 3 consisting of ILE (iron limitation at exponential phase), two replicates of IRE (iron repletion at exponential phase), and one replicate of GSNO (nitrosative stress). Lastly, cluster 2 consists of 5 % chicken juice replicates and two cold stress replicates to no surprise as it was established that there were no pathways enriched from the significantly differentially expressed genes between these two conditions, and this clusters together with cluster 1, which contains the all the remaining conditions. Most notably E37 (standard growth in MH2 at 37 °C exponential phase) did not cluster together with E42 (standard growth in MH2 at 42 °C exponential phase) but was closer to acid shock and the remaining cold shock replicate. E42 clustered more closely with bE and one replicate of bS (i.e. sodium deoxycholate exponential and early stationary respectively). Heat shock and anaerobic stress clustered closely together as did the two replicates of GSNO, one replicate of IRE, and two replicates of bS.

Figure 5.24 was divided into five sections according to the stringent genes for ease of interpretation. It is clear that the heat map is defined by two major clusters: sections I and II of the top half, and sections III, IV, and V of the bottom half. In the top major cluster, section II consists entirely of ncRNAs that were manually selected in Table 5.10. There were two ncRNAs in section II: as_Cj1270c and as_Cj0274_1 that had no expression ($<$ 10 TPM) across the 63 replicates (excluding chicken juice) as highlighted by the purple box. But the genes antisense to these ncRNAs Cj1270 and Cj0273 (*fabZ*) had expression across all conditions. The majority of antisense ncRNAs in section II were not expressed in many conditions or had low TPM values when expressed. Although it seems many were expressed during osmotic stress. Section I is defined by genes which have relatively low TPM values but are expressed across all conditions, but less so in clusters 4 and 5. Within section I there are 7 genes and 2 ncRNAs. as_Cj1140_1 is antisense to *cstIII* which is found in section V and has slightly higher expression but across all the conditions with subtle variations. However, as_Cj1140_1 was identified by Dugar *et al.* (2013) and runs into gene Cj1139c so its expression may be from the overlapping gene region. The other ncRNA is as_Cj0353c which is a novel ncRNA identified from toRNAdo, the gene antisense Cj0353c is a *ppx/gppA* homologue and is also found in section V with low expression across all conditions. Strangely, the genes *purN* (Cj0187c) and Cj0188-9c are adjacent to each other in an operon but the preceding genes starting from Cj0193c cluster elsewhere. Cj0190c is in section V with low expression but across all conditions, and *def* (Cj0191c) is in section IV and has higher expression but varies across the conditions. *clpP* (Cj0192c) is one of the genes of interest as part of the bacterial stress response and has even higher expression in section III along with gene at the start of the operon *tig* (Cj0193). Therefore, it seems overall the expression of the operon dies down depending on the order of the gene.

In section III, the gene *tatA*, which encodes for a twin-arginine transport protein, has fairly high expression with some subtle variation in some replicates across all conditions suggesting that it is constitutively expressed as highlighted by the navy dark blue box, although the functional role of this hypothetical protein is unknown. It is the only highly expressed gene in cluster 4 and was identified as an essential gene in Mandal *et al.* (2017)'s study. This gene is directly downstream of *gmk* which was found to have interactions

with *spoT* from STRING, but the expression of *gmk* is slightly lower than *tatA* especially at L42 (late stationary 42 °C). Other highly expressed genes across all conditions apart from in cluster 4 (ILS) were the ribosomal RNA operon beginning with *rpmJ* and followed by *rpsM*, *rpsK*,and *rpsD*, *rpoA*, *rpoQ*, and finally *rplQ* as highlighted by the green box. *rpmJ* and the latter three encode for 50S ribosomal proteins, hence they cluster together more closely, and the rest encode for 30S ribosomal proteins. Strangely, *rpmJ* is coded in a different frame at the beginning of the operon preceding the genes encoding for 30S ribosomal proteins. There are two ncRNAs in section III: as_Cj0440c_2 and as_Cj0441c but again both ncRNAs are on top of the gene *acpP*, which is why they cluster together with *acpP* and *fabF* the adjacent gene downstream. *ppa* and *adk* had lower expression in L37 (37 °C late stationary phase) relative to the other conditions.

Section IV has genes with a more dynamic expression within each condition but expression across all conditions. A few genes have higher expression compared to other genes in the same cluster in some condition. For example, *pyk* which is predicted to interact with *spoT* from the STRING database has high expression in E42, cold stress, and cj5, lower expression in bE, osmotic stress, and starvation, very low in L37 and L42, and moderate in all other conditions. The two ncRNAs in this section identified from Dugar *et al.* (2013) are on top of the gene Cj1074c which it clusters with. The genes in this section seem to have lower expression in clusters 5 and 7 compared to the other clusters.

Section V contains the cluster of genes which have moderate to low TPM expression that are expressed in all conditions including *spoT*, which is highlighted in pink. There is relatively lower TPM expression of *spoT* in nutrient-deprived conditions of clusters 5, 7, and 1. The homologues of polyphosphate granules *ppk* and Cj0604 (*ppk2*) cluster together as well as Cj0605 even though *ppk* is encoded elsewhere on the genome as highlighted by the light blue box. In all the replicates for heat shock and two replicates of cold shock, Cj0604 had higher expression that other genes in this section. There are four ncRNAs: as_Cj0440c_1, as_Cj0022c, and as_Cj0023c which all cluster together and as_Cj1359_3 in this section. As they were all identified by Dugar *et al.* (2013) they all overlap or run into genes. as_Cj0022c and as_Cj0023c are on top of *purB* and run from the same promoter therefore the pattern of expression is very similar with only slight variations in some replicates and they all cluster together. As for as_Cj0440c_1, this ncRNA also runs into *acpP* but has a long UTR which may explain why it did not cluster with *acpP* and as_Cj0440c_2. Finally, as_Cj1359_3 runs into the gene *nrfH* which is not included in the list of stringent genes. However, as_Cj1359_3 does also have a long UTR that is opposite *ppk*. There was < 10 TPM of expression for the promoter tss4060_as_Cj1359_3, from which as_Cj1359_3 runs off. The ncRNA also overlaps with other identified antisense promoters which also had very low TPM expression around the cut-off, so it is possible that the expression is due to the overlapping gene region.

The $\log_2$ fold-change of stringent response genes of comparisons in Table 5.6 are shown in Figure 5.25. Only the individual pairwise comparisons were used and was divided into six sub-clusters denoted by the annotated dendrogram. The clustering was based on absolute $\log_2$ fold-change values (i.e. all negative values were converted to positive values) so that highly changing genes would cluster together regardless of up or down-regulation. All $\log_2$ fold-changes have a *p*-adjusted value ≤ 0.05 but no threshold was used for the $\log_2$ fold-change values. A set of ncRNAs had no differential expression in any of the 32 comparisons used, highlighted by the pink box. But surprisingly, some ncRNAs were the most highly changing of the stringent response gene list. A cluster of ncRNAs had very high fold-changes in sub-cluster 6, which has the comparison G-L/E-37 and sub-cluster 3, which has S-osmotic/E, I-37/IL-E, I-37/IL-a, C-cj5/E37, and C-4/E37 highlighted by the blue box. These are: as_Cj0187c_2 identified by Dugar *et al.* (2013) and as_Cj1178c_1, a novel ncRNA identified from toRNAdo in this study at the top. Two ncRNAs: as_Cj0230c and as_Cj0232c at the bottom in the yellow box were highly changing in sub-cluster 4 containing both I-S/E-IL and I-37/IL-S, which have an inverse pattern of $\log_2$ fold-change, and in I-IR/IL-S in sub-cluster 2. Both ncRNAs were identified by Dugar *et al.* (2013) and although they do not overlap one other, they are positioned successively from one another. as_Cj0230c is antisense to Cj0230c but also to *nrdB*, whereas as_Cj0232c is antisense to Cj0232c and *pyrE*.

Sub-cluster 4 also shows many ribosomal RNA genes that are moderately changing and most of the other genes to have some changes suggesting the stringent response is very active in these two comparisons. *tatA* had high expression across all conditions in Figure 5.24 but the $\log_2$ fold-change was more dynamic between the comparisons. *spoT* was not differentially expressed in around half of the comparisons and only had very low $\log_2$ fold-changes as highlighted by the purple box, which was unexpected given its deduced role in the stringent response.

Figure 5.25: Heat map of $\log_2$ fold-change of stringent response genes. Only values that were significant i.e. p-adjusted value ≤ 0.05 were included. Important genes and sections are annotated in coloured boxes and explained in detail in the main text.

5.6.1.1 TPM expression and fold-change of stringent response genes

SpoT is the key regulator of the stringent response in *C. jejuni* as it regulates production of (p)ppGpp which is thought to alter gene expression by binding RNA polymerase (Gaynor *et al.*, 2005). The expression of SpoT did not vary much across the conditions in terms of TPM expression. There were subtle variations in Figure 5.24 (highlighted by the pink box) where it is clear to see there was slightly higher expression within replicates of some conditions including E42, bE, anaerobic, bS, GSNO, acid, cold, E37, cj5, ILE, IRE, and S42. However, the remaining conditions including starvation and late stationary phase conditions had lower TPM expression. In terms of fold-change, in Figure 5.25 highlighted by the purple box, SpoT was up-regulated in C-cj5/E37, I-37/IR-E, I-37/IL-S, I-42/37-S, and I-37/IR-S; and down-regulated in I-S/E-IL, I-IR/IL-E, G-L/E-37, G-L/S-37, G-L/E-42, G-42/37-L, and S-osmotic/E37. It seems that SpoT is expressed more in at 37 °C exponential when compared to other conditions.

The exopolyphosphatase/guanosine pentaphosphate phosphohydrolases (ppx/ggpa) have potential dual function for poly-P hydrolyis and generation of (p)ppGpp (Malde *et al.*, 2014). In NCTC 11168, the genes encoding for ppx/gppa homologues Cj1237 and Cj0353c were shown to interact with *spoT* and homologues in the strain 81176 were shown to contribute to the (p)ppGpp pool (Malde *et al.*, 2014). Cj1237 and Cj0353c did not cluster together in terms of TPM values across all conditions (Figure 5.24) though both showed low expression values among all the replicates. Cj1237 is in section I in Figure 5.24 and is strangely clustered more closely with the ncRNA as_Cj0353c, whereas Cj0353c is in section V. They also do not cluster together in regards to fold-change across comparisons (Figure 5.25). Cj1237c is up-regulated albeit weakly in the following comparisons: G-S/E-42, G-42/37-S, I-S/E-IR, G-L/E-42, and I-S/E-IL so it is seems to be expressed more in the later growth phases at 42 °C and in early stationary in iron-limited conditions. Cj1237c was down-regulated in I-IR/IL-S, I-IR/IL-E, I-37/IL-S, and C-4/E37 suggesting it is up-regulated in MEMα. Cj0353c was up-regulated in B-b/37-S, B-b/37-E, S-anaerobic/37, G-42/37-L, I-37/IR-S, and I-37/IL-S. So it seems in the presence of bile this gene is more active, and at early stationary phase it is expressed more in rich media than minimal media.

Although indirectly linked to the stringent response through ppx/gppa and levels of (p)ppGpp, polyphosphate kinases were included in the stringent response gene list (Table 5.10). NCTC 11168 has both homologues to *ppk1* - Cj1359 named as *ppk* and *ppk2* - Cj0604 in the NCBI reference genome AL111168 AL139074-AL139079 (NCTC 11168). Both polyphosphate kinases cluster together in TPM expression and fold-change along with Cj0605, which is downstream of Cj0604 (homologue of *ppk2*), highlighted by the light blue box and brown box respectively. Interestingly, TPM expression of Cj0604 is slightly higher than *ppk* or its downstream gene Cj0605 in heat shock, cold stress, and iron repletion at early stationary phase (Figure 5.24). With regards to fold-change most of the patterns are the same for all three genes although there are some slight variations between the polyphosphate kinases. All three genes were up-regulated in B-b/37-S, B-S/E-b, I-S/E-IL, and I-37/IL-E indicating the enzymes are more active in the presence of bile at early stationary phase and down in exponential phase in the absence of iron. *ppk* was

only up-regulated in G-42/37-S and S-starvation/S37 whereas Cj0604 was only up-regulated in I-S/E-IR, C-4/E37, I-37/IL-E. All three genes were down-regulated in G-S/E-37, G-L/S-42, B-b/37-E, G-L/E-42, S-osmotic/E37, G-L/E-37, and for just the polyphosphate kinases G-L/S-37. *ppk* was only down-regulated in I-IR/IL-S, C-cj5/4, and C-cj5/E37 whereas Cj0604 was down-regulated in S-GSNO/E37 only. Production of poly-P also requires phosphates and the ABC transporter phosphate uptake system was enriched for some comparisons in Figure 5.21, Section 5.4.7. However, there was no clear pattern between phosphate uptake and expression of polyphosphate kinases. Both *ppk* and Cj0604 were down-regulated in G-L/E-37, G-L/E-42, G-L/S42, and S-osmotic/E37 and the ABC transporter uptake system was simultaneously enriched from both up and down-regulated genes for these comparisons. However, the same gene *pstS* which is a periplasmic phosphate binding protein was down for all four comparisons.

5.7 Discussion

HTS technologies have advanced rapidly in the past decade whilst the cost has reduced dramatically since their conception. Yet, the process from ribosomal RNA depletion, library preparation, to sequencing is still very expensive and not ideal if there are many samples adding to the costs. RNAtag-Seq is a method which can produce the same results as normal RNA-seq but at $1/10^{th}$ of the cost depending on the number of unique tags used. Here we have validated the RNAtag-Seq protocol devised by Shishkin *et al.* (2015) and shown that it can be used in conjunction with different bacterial species on the same cartridge.

Data obtained from RNAtag-Seq runs was used to explore the transcriptional changes during a series of growth conditions and stress shocks. The transcriptome of some of these stress conditions used for this project have not been studied before, namely chicken juice related conditions. This is also the first transcriptomic compendium to be carried out in *C. jejuni* to date. This data was used alongside data from Chapter 4 for identifying ncRNAs. As this data set can provide a valuable resource for the scientific community it is important to ensure that it is reproducible of studies in the literature for some conditions and of good quality.

5.7.1 Caveats and reproducibility of RNAtag-Seq

RNAtag-Seq was successfully used to sequence 66 samples over two sequencing runs but a major caveat was discovered. The sequence of the tags used which were 8 bp long and had an extra 'T' nucleotide was incorporated into Read 1 during library preparation (see Section 2.5.2). This was due to the orientation of the barcoded adapter placed downstream of the sequencing primer and was therefore sequenced as part of the 76 cycles (with one extra cycle for phasing) (see Figure 2.1 Chapter 2). Thus 9 bp of sequence information was lost from each forward read (read 1). This also caused complications with the automated de-multiplexing on Illumina Basespace, although this was overcome by writing customised scripts to de-multiplex and trim the raw fastq files. As we had good phred quality (Q) scores of 30 (Q30) from our sequencing run, no mismatches were allowed when de-multiplexing the tag sequences contrary to the 1 bp mismatch parameter that Shishkin *et al.* (2015) allowed in their study. A QC30 score indicates 1 in 1000 probability of an incorrect basecall or 99.9 % basecall accuracy, so for every 1000 bp there may be an error. The fragment size used for preparing RNAtag-Seq libraries was confirmed from bioanalyzer results (data not shown) to be $\approx$500 bp so it is highly unlikely there was an error during sequencing, hence any tags with a 1 bp mismatch would not be trusted as usable data.

All samples from the sequencing runs except for the chicken/cold-related conditions had over 90 % alignment to the reference genome (Table 5.1), which is slightly lower than usual for bacterial genomes but still successfully generated read counts for DESeq2 input, verifying the reproduceability of RNAtag-Seq. Chicken juice would be expected to contain cells from the chicken and this was confirmed when aligning reads to the *Gallus gallus* genome strain GRCg6a with a big percentage of reads from 100 % chicken juice, which explains the low alignment to NCTC 11168. However, it was surprising that cold stress also

has low alignment rates with the lowest replicate having 68.87 % alignment to the NCTC 11168 genome. As these samples were treated the same way as all other samples, it is unlikely that the low alignment was due to mishandling during preparation, although this cannot be ruled out. It may be more plausible that the nature of the stress (24 hours at 4 °C) may be causing more RNA degradation as compared to the other stress conditions. Although the RNA degradation pathway was not enriched in the cold-related conditions (Figure5.18), we cannot rule out the possibility of this happening in the earlier stages of the stress. *C. jejuni* does not possess the typical cold shock proteins such as the RNA chaperone CspA in *E. coli* among others which help stabilise secondary structures of nucleic acids affected by the low temperatures indirectly facilitating transcription, although it does have PNPase but the lack of cold shock proteins may be a reason for the poor RNA quality obtained (Phadtare and Severinov, 2010). There was a significant drop in CFU counts after applying cold stress, which was also the case for 5 % chicken juice and many other stresses. Subsequently, only the mapped reads of good quality were used for further analysis.

It is important to study the variation of the raw data and ensure replicates cluster together. The first sequencing run had one replicate which was an outlier (Figure 5.1) even with a high alignment rate to NCTC 11168 and similar number of reads (Table 5.1). This could have arisen at any of the RNA extraction, library preparation, or sequencing stages. Unfortunately, the outlier was from 37 °C exponential which was used as the control group to compare many conditions against, so another replicate was sequenced on the next run. Iron repletion early stationary was also repeated as the MEMα media used for this condition had a colour change (from red to pink) indicating a change in pH due to oxidation (i.e. more acidic). To prevent any batch effects, it was decided to repeat this condition with a new bottle of MEMα, which allowed a more comparable experimental design with the other iron-related conditions.

As seen in Figures 5.2 and 5.3, biological replicates from different runs cluster together as well as clustering in the same run. This implies that biological replicates from different days from a different experiment, library preparation, and sequencing run do not contribute to the variation shown in gene expression for this data set. Only the replicates from the second run for 37 °C exponential and iron repletion early stationary were subsequently used for further analysis. Figure 5.4 shows all biological replicates from each condition cluster well together further supporting the robustness of the data set and good reproduceability between batches of samples. This was also supported by the clustering seen in Figure 5.24 where most biological replicates clustered together. PC1 in Figure 5.4 has a maximum variation of 36% with nutrient deprived conditions clustering together in the red circle on the left. This means that nutrient depletion and late stationary phase show the biggest variation between the conditions, although 36 % is still not very high. However, sodium deoxycholate early stationary is not part of this cluster and seems to be more similar to sodium deoxycholate exponential, which has the same growth supplement but different phase, whereas iron repletion early stationary is surprisingly part of another cluster circled in green at the bottom with osmotic stress and 100 % chicken juice. This implies that the growth phases act differently under different stresses.

Genes that were found to be up or down-regulated in certain conditions correlated with what was found in the literature to indicate the robustness and quality of the data set. However, there were some discrepancies observed. CjNC19 is an intergenic ncRNA found by Porcelli *et al.* (2013) and in a previous study conducted in the lab by Stoakes (2017), CjNC19 was found to be one of the top ten most down-regulated gene features in a pseudorevertant of a deleted *flhF* NCTC 11168 strain. In this study, there was no TSS identified for this ncRNA and although the ncRNA was included in the GFF file there were no reads mapping to this region in all conditions.

5.7.2 The enrichment analysis of grouped and individual pair-wise comparisons from DESeq2

As the transcriptional response to a compendium of conditions were sequenced, this allows for a vast number of comparisons between conditions and grouped comparisons using two-factor analysis to determine whether a factor is contributing to the differentially changing genes. However, a big data set is difficult to analyse and interpret because of the large amount of data and the vast possible different pairwise combinations with 22 different conditions. A simplified approach is to group the conditions together and decide which comparisons to make. The most logical comparisons were made by assuming 37 °C exponential as the wild type 'default' especially for stress shocks as these were applied at exponential phase. However, for conditions where an addition of supplement or change in media was used from the very beginning for example when comparing iron-related conditions, iron limitation was used as the control as the defined media MEMα does not contain any iron, whereas the constituents of MH2 contain traces of iron, and iron repletion adds $FeSO_4$ as a supplement.

By analysing all 22 conditions together, common patterns and trends can be easily identified with some conditions exhibiting similar behaviour to each other, whereas others may have unique responses. It was surprising that osmotic stress clustered separately from other stresses. But it is more similar to late stationary stresses and 100 % chicken juice. 1.5 % NaCl was used for osmotic stress and the resulting osmolarity is $\approx$0.8 osmol liter $^{-1}$, which is in between chicken caecum at 0.7 osmol liter $^{-1}$ and the chicken duodenum at 0.9 osmol liter $^{-1}$, so this may have explained why similar pathways were deactivated with 100 % chicken juice as there could be components in both conditions which trigger *C. jejuni* to respond as it would in a chicken host (Cameron *et al.*, 2012). Although given that 100 % chicken juice did not have good RNA samples this may be just an artefact. Nevertheless, osmotic stress was still clustered separately from all other conditions and behaving in an unusual way.

Under some comparisons the Ribosome pathway was enriched. The down-regulation of ribosomal protein genes during stress is likely due to the need of delegating energy towards expressing stress resistance genes as a priority rather than energy-consuming translation, as was seen previously in an experiment during acid shock (Reid *et al.*, 2008). In our data, acid stress did not enrich for the Ribosome pathway but

strangely other stresses such as starvation, anaerobic, heat, and nitrosative stress did (Figure 5.19). The only comparisons which down-regulated the Ribosome pathway were mainly iron-related comparisons, cold-related comparisons but with very weak *p*-adjusted values, and one bile comparison B-S/E-a.

The other two comparisons which strongly enriched for the Ribosome pathway from up-regulated genes (in red) was the iron-related comparisons I-37/IL-S and I-37/IR-S (Figure 5.16). Both comparisons have growth in MEMα as the reference indicating that at early stationary phase, ribosomal protein genes are down-regulated in defined media as compared to the complex nutrient rich MH2 broth. It is possible that the stress during early stationary phase in defined media is more severe than in rich media. This was also demonstrated by the enriched Ribosome pathway from down-regulated genes (in blue) of I-S/E-IL and I-S/E-IR which are both early stationary against exponential phase in MEMα but with the absence and presence of iron respectively indicating that early stationary phase in minimal media is more severe than exponential phase in minimal media. In terms of host temperature the Ribosome pathway was strongly enriched from down-regulated genes in G-42/37-S and weakly enriched by up-regulated genes in G-L/E-42 (Figure 5.11). Although the pathway was not enriched in G-42/37-a, which removes the effect of growth phase and only looks at the difference between host temperatures. This suggests that between the host temperatures at early stationary there is a stark difference in expression of ribosomes. The bile and cold-related comparisons did not enrich for the Ribosome pathway.

5.7.2.1 Difficulties with chicken juice samples

Studies have investigated the enhanced survivability of *C. jejuni* at 4 °C in the presence of chicken juice but this was the first time the transcriptional profile has been investigated during cold stress, and in the presence of chicken juice (Birk *et al.*, 2004; Brown *et al.*, 2014; Karki *et al.*, 2019). Figure 5.18 shows most significantly enriched pathways to be inactivated in 100 % chicken juice with the exception of ABC transporters implying arrested growth. Although *C. jejuni* cannot replicate at 4 °C but chicken juice does contain nutrients to support growth at host temperatures (Birk *et al.*, 2004). Figure 3.19 in Chapter 3 showed that 100 % chicken juice did not have a significant drop in CFU after 24 hour incubation at 4 °C as compared to 5 % chicken juice and cold stress implying a protective component. Birk *et al.* (2004) also found that *C. jejuni* viability was prolonged in chicken juice as compared to BHI broth and BHI supplemented with 5 % calf blood at 4 °C over a 30 day period, with even greater impact at elevated temperatures. This also corroborates with Karki *et al.* (2019)'s study which also tested chicken liver juice and found the additional components from liver enhanced survivability even further (Karki *et al.*, 2019). Therefore, it is surprising that many KEGG Metabolic pathways were strongly down-regulated in 100 % chicken juice.

It is important to take into consideration that there was a low percentage of mapped reads in 100 % chicken juice samples as seen in Table 5.1, with a majority of reads actually mapping to the chicken genome (Table 5.2). This could explain why when compared with 37 °C there was a high number of significantly enriched inactivated KEGG pathways. The RIN values for chicken juice were $\approx$2 (data not shown) which

is very low. The difficulty in extracting RNA could possibly be from increased biofilm production whilst in the presence of chicken juice as was shown in Brown *et al.* (2014). Due to time constraints the RNA extraction for chicken juice was never optimised and thus, the resulting RNA-seq data is not dependable, although it can form a basis for future optimisation such as deeper sequencing and/or enriching for bacterial RNA from biofilms. Since the low quality of chicken juice samples were not expected, the DESeq2 pipeline included this sample for normalisation. Due to time constraints the analysis was not repeated without the samples. Although this would only have a minor effect on some of the results, it is highly advisable to remove the samples and possibly redo the analysis in the future for more accurate results.

5.7.2.2 Growth phase has a higher impact on metabolic response in *C. jejuni* than host temperature

Many KEGG metabolic pathways were strongly activated or inactivated when compared against different growth phases, but at the same growth phase and different host temperature there were not as many pathways enriched. This suggests that the major changes in transcription are due to growth phase more than temperature. Perhaps the changes in host temperature which determine *C. jejuni* as a commensal or pathogen are more subtle. In particular, at the differences between host temperatures at late stationary were more pronounced. In general, there were more metabolic pathways that were up-regulated at 42 °C.

5.7.2.3 Flagellar assembly across different stress conditions

Flagella is important not just for virulence but also for survival and motility in *C. jejuni* and aflagellate mutants have been shown to have attenuated colonisation in chicks (Nachamkin *et al.*, 1993). They may also be useful for coping with stresses such as acidic environments to aid in its agility to move away from undesirable conditions (Roure *et al.*, 2012). Our data clearly shows that Flagellar assembly is strongly activated during early and late stationary phases at both 37 °C and 42 °C (Figure 5.11). However, when 42 °C was compared with 37 °C at late stationary phase (G-42/37-L) many down-regulated genes enriched for the Flagellar assembly pathway. In addition, σ^{54} was down-regulated in G-L/E-42, which corresponds to mid flagellar genes down-regulated in Figure 5.14 in the same comparison. Our results suggest that host body temperature does have an effect on flagellar biogenesis but only during late stationary phase and not as pronounced as the different growth stages. This is in contrast with Aroori *et al.* (2013)'s study which implied an induction of *flaB* expression at 42 °C over 37 °C, although this was not statistically significant and different strains were used. Since flagella is a known virulence factor and it is suggested that *C. jejuni* may depend on temperature to sense whether it is in a chicken or human host, it would be expected that at 37 °C where *C. jejuni* is more pathogenic, that flagellar genes would be expressed more (Aroori *et al.*, 2013).

Wösten *et al.* (2010a) also implied that the interaction between σ^{28}/FliA and FlgM, an anti-σ^{28} factor was stronger at 37 °C and inhibits the FlgRS two-component system which regulates σ^{54}-dependent genes. Meaning at higher temperatures i.e. 42 °C there is an up-regulation of σ^{54}-dependent genes due to less σ^{28}/FliA-FlgM interaction (Wösten *et al.*, 2010a). This contradicts with our data where there was

down-regulation of σ^{54}-dependent genes (and σ^{28}) at 42 °C compared with 37 °C at late stationary phase (Figure 5.12). However, Wösten *et al.* (2010a) did not find the σ^{28} or σ^{54} promoters to be temperature-dependent by qPCR, so it is possible that the effect seen is from the activity of the protein but not the expression of the protein. Our data shows a slight negative fold-change at 42 °C compared to 37 °C at both early and late stationary phase $>$ -1 $\log_2$ fold-change but a 5.71 $\log_2$ fold-change at exponential phase. This is in contrast with the 2 fold-change observed by Wösten *et al.* (2010a) although this may be due to differences between qPCR and RNA-seq, and the growth phase used. FlgM was also in the list of highly expressed genes from the 10th decile of TPM normalised values across all conditions (Table 5.8). Further investigation is required to understand the intricacies of FlgM.

It is not surprising that flagellar biogenesis is activated during the later growth phases as flagella have hierarchical assembly, and this was also seen in the transition from exponential to early stationary growth in MEMα under the presence and absence of iron (Figure 5.17). Many flagellar genes are known to be iron-regulated, and iron-limitation could be an environmental cue that *C. jejuni* is inside a host. As MH2 broth only has traces of iron as explained in the next section, it is possible that the depletion of iron in early and late stationary phase is also contributing to the increased expression of flagella. Strangely, during the transition from exponential to early stationary phase in sodium deoxycholate, the KEGG Flagellar assembly pathway was not enriched. However, the Flagellar assembly pathway was also enriched under anaerobic stress even though the stress was applied during exponential phase. It is possible that anaerobic stress is also a cue for a particular location in the host, for example, the caecum, and since *C. jejuni* prefers microaerophilic conditions, the rapid increase in flagellar biogenesis will allow the bacteria to swim towards more favourable conditions.

Flagellar assembly was not enriched during acid shock in our data. The relationship between acid shock and flagellar expression is controversial as there are conflicting reports of flagellar genes up-regulated or down-regulated in acidic conditions in other bacteria (Varsaki *et al.*, 2015). However, two studies carried out in *C. jejuni*: Varsaki *et al.* (2015) and Le *et al.* (2012) found flagellar genes to be up-regulated in response to acid shock. Varsaki *et al.* (2015) only reported three flagellar genes *fliGH* and *maf3* that were up-regulated in acidic conditions at 42 °C but used strain CI 120, which is known to have an adaptive acid tolerance response not found in other strains. Le *et al.* (2012) reported a subset of σ^{54} flagellar genes in the *C. jejuni* strain NCTC 11168 were up-regulated during acid shock at pH 5 and pH 3.6 in Brucella broth for 10 minutes at 37 °C using microarrays. Our data used similar conditions at pH $\approx$3.5 but in MH2 broth and did not find any flagellar genes that were significantly up-regulated even without a fold-change cut-off.

5.7.2.4 ABC transporter networks are subject to environmental change

The uptake of many substrates, ions, small and macro-molecules except a minority of some elements which are able to freely diffuse across the bacterial cell membrane are aided by many transporters, the largest family consisting of the ATP-binding cassette (ABC) (Rees *et al.*, 2009). ABC transporters are present in all domains of life and in bacteria they can either import substrates for acquisition of nutrients or export toxic

molecules conferring antimicrobial resistance (Locher, 2016). ABC transporters were consistently enriched across the compendium of conditions used in this project demonstrating the importance and necessity of these transporters. Most notably many transporters were involved in the uptake of essential metals and in particular the Iron(III) ABC transporter was the most frequently occurring and was surprisingly involved in many of the conditions/comparisons in the group of stress shocks. Other metals uptake systems were also enriched such as Molybdate, Tungstate, and Zinc. The uptake of all these trace elements have to be highly regulated since in excess they are toxic to cells and insufficient amounts will arrest growth.

NCTC 11168 has five different iron uptake systems as described in the review by Miller *et al.* (2009). Though not the only ABC-dependent iron transporter, the Iron(III) uptake system from the KEGG pathway is the characterised transferrin/lactoferrin uptake system encoded by the *cfbpABC* operon. This uptake system typically transports ferric iron (Fe^{3+}) bound to transferrin found in host serum and bound to lactoferrin present in milk (Miller *et al.*, 2008). As expected, *cfbpABC* was significantly down-regulated in iron-repleted conditions to prevent the accumulation of excess iron inside the cell which can be toxic due to the production of ROS (Figure 5.21). This also indicates it is highly up-regulated in iron-limited conditions i.e. absence of iron. Intriguingly, *cfbpABC* was also down-regulated in I-37/IL-E, G-L/S-37, and G-L/S-42. The down-regulation of *cfbpABC* in I-37/IL-E may be down to the media used as standard growth was carried out in MH2 broth and iron-related conditions in MEMα, though at early stationary phase only one gene made the cut-off. The down-regulation at both host temperatures in transition from early stationary to late stationary phase could be due to the up-regulation of Iron(III) during transition from exponential to early stationary phase. This may be due to the regulation of iron levels, where during early stationary phase adequate uptake of iron was achieved, therefore the transporter was down-regulated in late stationary phase. It is also possible that at late stationary phase many metabolic pathways are shutting down as was seen in Figure 5.11, so there is no need for the uptake of iron if enzymes requiring iron as a co-factor are also not expressed. Cj0176c is the first gene at the start of the operon preceding *cfbpABC* and it was shown in Chapter 4 Section 4.3.2 Table 4.6 that the secondary novel promoter tss530_Cj0175c_2 for the whole operon had a holo-Fur activation sequence. This implies the presence of ferrous iron when bound to Fur potentially activates transcription of the Iron(III) ABC transporter network, although further studies are required to confirm this. But the data from DESeq2 analysis in Figure 5.21 does not support this, as the Iron(III) ABC transporter network was not activated in the presence of iron.

The Iron(III) uptake system was significantly up-regulated in transition from exponential phase to early stationary phase at both 37 °C and 42 °C and both combined temperatures. The MH2 cation-adjusted broth (Sigma-Aldrich) used in this project is manufactured by Merck and contains 17.5 g/L casein acid hydrolysate, 2 g/L beef extract, and 1.5 g/L starch adjusted to pH 7.4 $\pm$ 0.2 (Nussbaumer-Pröll *et al.*, 2019). Beef extract could potentially be a source of iron and a study by Girardello *et al.* (2012a) tested for metal ion concentrations of several brands of MH agar including Merck revealing the concentration to be 0.6 mg/L. The study was conducted in 2012 and since then the composition of the media may have changed so it is unclear what the current iron concentration is in MH2 (Sigma-Aldrich). Judging from

our data it seems that at early stationary phase the supply of iron may already have been depleted as the Iron(III) network was also up-regulated at I-37/IR-S, which is MH2 broth against MEMα supplemented with iron. It was also very surprising that the Iron(III) uptake system was up-regulated in all the stress shocks apart from osmotic stress and starvation. An excess of iron is toxic to cells due to the generation of ROS in conjunction with oxygen so it was unusual to increase the expression and uptake of ferric iron during peroxide stress, although paradoxically this was also the case for anaerobic stress. It is possible that the response in the uptake of iron may be linked to a general stress response which may require enzymes in pathways, which need iron as a co-factor.

It was interesting to see that a few comparisons enriched for the Molybdate and Tungsten ABC transporters for the uptake of the trace elements molybdenum and tungsten, as not much is known about these metal uptake systems. Whilst traces of molybdenum are ubiquitously present in the natural environment including seawater and soil and are widely used by all domains of life, tungsten is comparatively scarce and its use is restricted to prokaryotes (Smart *et al.*, 2009; Taveirne *et al.*, 2009; Smedley and Kinniburgh, 2017). Molybdenum and tungstate have similar chemical properties and sizes so it is difficult for the molydenum-binding protein ModA to distinguish the two, and it was thought that tungsten acts as an antagonist (Makdessi *et al.*, 2001; Smart *et al.*, 2009). However, in NCTC 11168 it was reported that Cj1540, the homologue of the tungsten-binding protein TupA and part of the Tungsten ABC transporter gene set, had very high affinity for tungsten as compared to molybdenum (Smart *et al.*, 2009). In addition, insertional inactivation of Cj1540 had $\approx$50 % reduction in formate dehydrogenase activity which oxidises formate and acts as an electron donor (Smart *et al.*, 2009). Formate dehydrogenase is one of four enzymes that utilise a metal binding pterin (MPT) co-factor with either molybdate or tungsten, and is involved in the electron transport chain (Taveirne *et al.*, 2009). Our data show that all three genes of the tungstate ABC transporter were significantly up-regulated at early stationary in the presence of bile (B-S/E-b) and a number of other comparisons also had one or two tungsten uptake genes up-regulated. The full set was also down-regulated at exponential phase in the presence of bile compared to no bile (B-b/37-E) and during osmotic stress (S-osmotic/E37) as seen in Figure 5.21. Therefore, it seems during exponential phase in the presence of bile, tungsten uptake is repressed. A study by Negretti *et al.* (2017) which performed RNA-seq with the addition of sodium deoxycholate in NCTC 11168, did not report this finding, although many factors such as the concentration of sodium deoxycholate, growth phase, and volume used were different to our experimental design. However, Negretti *et al.* (2017) did find many down-regulated genes that encode for the components of the electron transport chain (Negretti *et al.*, 2017). It is possible that the repression of tungsten indirectly affects the electron transport chain via formate dehydrogenase.

The entire Molybdate ABC transporter gene set was down-regulated in B-b/37-E, B-b/37-a, during nitrosative stress (S-GSNO/E37), and osmotic stress. Though there were no comparisons which had all three Molybdate uptake genes up-regulated, I-37/IL-S had two genes up-regulated and various other comparisons had one gene. Two other genes are potentially associated with molybdenum uptake: Cj0302c, which is an unknown protein but situated in the *modABC* operon between *modA* and *modB*, and the *modE* homologue

Cj1507c which was shown to repress the Molybdate and Tungsten ABC transport systems (Taveirne *et al.*, 2009). As expected, Cj0302c was also significantly down-regulated with $<$ -1 $\log_2$ fold-change for S-GSNO/E37 and B-b/37-E, although for G-L/S-37 there was only -0.6372 $\log_2$ fold-change. B-b/37-S also had -2.19 $\log_2$ fold-change for Cj0302c but in I-S/E-IR this gene was not significantly changed. Due to the location of Cj0302c, it is not surprising it behaves similarly to the down-regulated *modABC* operon in these comparisons, although whether Cj0302c plays a role in molybdenum uptake is unclear. What was unforeseen was that expression of the ModE homologue Cj1507c was down-regulated in B-b/37-E with -1.636 $\log_2$ fold-change and B-b/37-S had -1.753 $\log_2$ fold-change but not highly changing in other comparisons, implying that in these comparisons the down-regulation of these uptake systems was not due to the influence of Cj1507c. Although Cj1507c does not have a metal-binding domain, it represses the Molybdate ABC transporter in the presence of excess molydenum and tungstate and Tungsten ABC transporter in the presence of tungstate only (Stahl *et al.*, 2012). None of the *in vitro* growth and stress conditions we applied had an abundance of both trace elements, so the down-regulation of the uptake systems could be due to another factor at play.

Just like the other trace metals iron, molybdate, and tungsten, zinc is also an essential co-factor for many crucial enzymes in bacteria and is also severely restricted in the host as part of nutritional immunity (Mikhaylina *et al.*, 2018). As zinc is able to form more stable complexes than other metals, excess levels of zinc can compete and incorrectly bind to enzymes and is therefore employed by macrophages to kill pathogens inside phagosomes (Davis *et al.*, 2009; Mikhaylina *et al.*, 2018). Therefore, zinc homeostasis also has to be tightly regulated. The entire zinc uptake system was up-regulated in C-cj5/E37, and two genes in anaerobic stress, G-42/37-L, B-b/37-S, I-S/E-IL, nitrosative stress, and B-S/E-b. Whereas the entire zinc uptake system was down-regulated in G-L/S-42 and G-L/E-37 and two genes were down-regulated in osmotic stress and G-L/E-37. Many other comparisons also had one gene that was down-regulated. Up-regulation of the zinc uptake system in 5 % chicken juice may be due to higher levels of zinc present in comparison with the other media used, as chicken meat is a moderate source of zinc for humans (Ma and Betts, 2000). One explanation is that zinc may be vital for chicken colonisation and perhaps some unknown factor in 5 % chicken juice is perceived by *C. jejuni* as being inside the chicken host (Davis *et al.*, 2009). The down-regulation in late stationary comparisons could be attributed to the shutting down of many metabolic pathways and enzymes that require zinc as a co-factor.

The Aspartate/Glutamate/Glutamine ABC transporter network was enriched from up-regulated genes in comparisons in the presence of iron (Figure 5.21). The genes in the pathway are Cj0919c, Cj0920c, *peb1A*, and *pebC*. *peb1A* may have dual functionality as an adhesin molecule and has been established as important for colonising the chicken gut, although experimental studies found conflicting results (Rubinchik *et al.*, 2012). However, inactivation of *peb1A* resulted in impaired growth in MEMα supplemented with aspartate and glutamate, and *peb1A* was shown to bind to both L isoforms of aspartate and glutamate emphasising its importance in the ABC uptake system (Del Rocio Leon-Kempis *et al.*, 2006). It has already been accepted that *C. jejuni* rely heavily on amino acids as a carbon source. Aspartate is one of the first amino acid prefer-

ences of *C. jejuni* after serine, and glutamate is also widely used, therefore there are redundant transporters in place for the uptake of these amino acids (Hofreuter, 2014). The full set of genes was down-regulated in B-b/37-E but up-regulated in B-S/E-b. The uptake system was also partially enriched in I-37/IL-E and I-37/IL-S, and fully enriched in I-IR/IL-S and I-IR/IL-a from up-regulated genes suggesting the uptake system is active under nutrient rich conditions, although this does not explain why it is up-regulated in B-S/E-b. Under hyperosmotic stress in the *C. jejuni* strain 81-176, *peb1A* was down-regulated 4-fold after 12 hour incubation in 2 % NaCl. Our data shows *peb1A* had -2.32 $\log_2$ fold-change after 2 hours in 1.5 % NaCl and the other genes in the operon also having $\leq$ -1 $\log_2$ fold-change, but only Cj0919c and Cj0920c significantly enriched for the uptake system (Figure 5.21, Section 5.4.7).

5.7.3 Highly expressed and highly variable genes

The highest changing gene from DESeq2 analysis was Cj1384c, which has been shown by published studies and our data to be highly repressed by iron. Novel putative promoters for this gene have been identified from our Cappable-seq data as seen in Table 4.5 of Chapter 4. Interestingly, this gene is also highly changing between other comparisons such as up-regulated 13.42 in late stationary against exponential phase at 37 °C (G-L/E-37) $\log_2$ fold and TPM expression analysis also corresponds with the DESeq2 data. This gene could be a candidate for potential future study. TPM normalisation was carried out for all genes to normalise gene length and sequencing depth within samples, and also partitioned into deciles to categorise highly expressed and conditionally expressed genes. Surprisingly, from the 61 highly expressed genes, only nine were essential (Table 5.8). These are genes that are constitutively and highly expressed within each sample/condition but not comparable with other conditions. It is important to take into account that the distribution of expression values is different for each condition so these genes can still be differentially expressed across different conditions. Although some genes are expected to be highly expressed such as the well known *porA* gene, other genes and ncRNAs have unknown function. However, it has been established that the highly expressed ncRNAs overlap with genes and their authenticity is questionable.

5.7.4 Osmotic stress exhibits unusual transcriptional behaviour

From the PCA plot without ncRNAs (Figure 5.4), it was shown that osmotic stress was close to 100 % chicken juice and iron repletion early stationary separate from the other conditions, so it was no surprise that osmotic stress clustered more closely with 100 % chicken juice comparisons in Figure 5.10. Although there were many metabolic pathways which were significantly deactivated, there were also some amino acid related pathways which were activated although not very strongly. This is strange since during stress bacteria tend to redirect their energy sources away from biosynthesis and growth. In particular, tryptophan biosynthesis was significantly enriched in this study and also in strain 81-176 in Cameron *et al.* (2012)'s study, even though it is the most energy consuming amino acid to be synthesised (Priya *et al.*, 2014). During osmotic stress, cell integrity is majorly affected due to water leaving the cell as a result of osmosis, leading to cell shrinkage or turgor where there is reduced hydrostatic pressure

against the cell membrane (Cameron *et al.*, 2012). In *Saccharomyces cerevisiae*, tryptophan and tyrosine was shown to be necessary for protection from membrane disruption due to SDS stress (Schroeder *et al.*, 2018). The Phenylalanine, tyrosine, and tryptophan biosynthesis KEGG pathway was up-regulated in osmotic stress implying these amino acids may also be playing a role in membrane protection in *C. jejuni*.

The top ten list of up-regulated and down-regulated genes suggested that tRNA-methyltransferase was negatively regulated during osmotic stress (Table 5.7). tRNA-methyltransferases are important for tRNA modifications and subsequent downstream protein synthesis, as unmodified tRNAs could lead to mis-translation (Hori, 2014). Although some amino acid synthesis pathways were activated, Cysteine and methionine metabolism was down-regulated (Table 5.7). Normally protein synthesis is repressed under stress so that precious energy and resources can be diverted to expression of genes, which help to cope with the stress, thus there is no need to maintain tRNA modifications. Many genes related to possible periplasmic proteins were also down-regulated. As osmotic stress affects cell membrane integrity due to the efflux of water, it is reasonable that expression of membrane proteins would go down.

There was a high number of differentially expressed ncRNAs in osmotic stress that were up-regulated. Moreover, in Chapter 4 there were a number of novel putative ncRNAs identified in osmotic stress only. However, this could be mainly due to pervasive transcription which could be elevated under harsh stress conditions. Indeed, the RNA degradation pathway was enriched from up-regulated genes in osmotic stress compared to standard growth conditions (Figure 5.19) by *dnaK*, *rho*, and Cj1710c. *rho* is a transcription termination factor that stops transcription elongation so its up-regulation may be attributed to an increase in pervasive transcription.

5.7.5 The stringent response exhibits subtle changes in nutrient limited conditions

The stringent response is known to be triggered by nutrient limitation so it was surprising that many of the key genes involved in the stringent response such as *spoT* and *ppk* were moderately expressed during exponential phase, whereas the expression was low at late stationary phase and nutrient deprived conditions. Although *spoT* is bi-functional, and is able to synthesise and degrade (p)ppGpp, it is possible that expression is low during stress to reduce degradation of (p)ppGpp. It is also strange that favourable conditions such as 37 °C exponential phase (E37) clusters with other stress conditions and have similar expression patterns. What is striking is that ILS (iron limitation early stationary) clusters separately from all other conditions and the stringent response seems to behave very differently in this condition. Further studies will have to be conducted to explore this in more detail. It was also surprising to see many ribosomal protein genes that were highly expressed, even in clusters 5 and 7, which contain all the nutrient-limited conditions apart from cluster 5 with the ILS replicates.

SpoT was found to have slightly higher expression in standard growth at 37 °C exponential phase in MH2 media. Although it was expected for SpoT to perhaps behave differently and become more expressed in stresses in particular relating to nutrient deprivation, one explanation for its moderate expression at

37 °C exponential phase and down-regulation in late stationary phase conditions could be due to the bi-functionality of SpoT. As SpoT is able to hydrolyse and synthesise (p)ppGpp, perhaps it is crucial in nutrient-rich environments to ensure (p)ppGpp does not accumulate, as binding of (p)ppGpp to RNA polymerase will prevent transcription from occurring which is more detrimental to the cell (Starosta *et al.*, 2014). This may also explain why *neuA1, cgtA* has little to no TPM expression in sub-clusters 4, 7, and 5, which are all nutrient deprived conditions (Figure 5.24) as it was speculated that when ppGpp reaches a high concentration, CgtA is able to bind to ppGpp and act as a 50S anti-association factor preventing the 50S ribosome subunit associating with the 30S ribosome, hence inhibiting translation, although this activity is yet to be determined in *C. jejuni* (Ronneau and Hallez, 2019). It may be possible that our stresses simply did not activate the stringent response, though this is surprising as the stringent response would normally be activated in nutrient deprived conditions such as early and late stationary phase or starvation, which we do not see in our data.

Polyphosphate kinases control the production of poly-P which play a role in adaptation of stress and induces formation of VBNC cells, with *ppk1* forming poly-P from ATP and *ppk2* forming poly-P from GTP (Gangaiah *et al.*, 2009). As poly-P are implicated as having a role in stress linked to (p)ppGpp levels, it was surprising that during late stationary comparisons and osmotic stress, both *ppk1* along with *ppk2* were significantly down-regulated. These comparisons also had the phosphate uptake gene *pstS* down-regulated, although some also had one or two phosphate uptake genes that were up-regulated. However, poly-P synthesis is reversible, so it is possible that down-regulation of this gene is to prevent lowering the concentration of poly-P that is present, even though poly-P synthesis is favoured (Malde *et al.*, 2014). Although poly-P degradation can also be attributed to ppx/gppa, which contributes to ppGpp levels in *C. jejuni* (Malde *et al.*, 2014). Both ppx/gppa seem to be expressed more in the late stationary phases at 42 °C and in MEMα, suggesting that in nutrient deprived conditions the poly-P levels were decreased. Polyphosphate granules, which are reserves of poly-P and act like an energy storage, have been observed under the electron microscope for *C. jejuni* (Müller *et al.*, 2014). Some bacteria seem to produce polyphoshate granules constitutively while others seem to only produce them under nutrient stress (Racki *et al.*, 2017). Although it is unknown which category *C. jejuni* is under, members of our laboratory have seen similar observations to polyphosphate granules at the polar ends of *C. jejuni* cells retrieved from early stationary phase under the electron microscope. It may be possible that both polyphoshate kinases are more active at exponential phase due to the production of poly-P, but once it is accumulated and stored in granules there may no longer be a need for such production, especially if it is energy consuming. Racki *et al.* (2017) suggests that polyphosphate granule biogenesis during nitrogen limitation in *P. aeruginosa* may be due to poly-P acting like a phosphate sink. It is not clear if this is occuring in our data and further investigation into the expression of phosphate uptake genes is needed.

5.7.6 Summary

With such an extensive set of data from a collection of 22 conditions and given the time constraints within this project, it was not possible to explore and analyse every aspect of the data. However, the preliminary analysis carried out in this chapter will hopefully set a foundation for future studies and projects for years to come, and provide a useful database for the *Campylobacter* research community along with the primary transcriptome in Chapter 4. There were also many contradictions between what was found in the literature and our data set, although it is worth noting that for some particular conditions our data merely looks at adaptation to stress, so a transient response rather than a maintained survival response to the stress may explain the discrepancies observed.

One hurdle of analysing this data was that often genes enriched for multiple pathways and there were redundancies in some pathways such as the general KEGG Metabolic pathways, which encompasses all metabolic pathways. For easier interpretation it may be useful to use a tool such as EnrichmentMap to improve visualisation of pathways (Reimand *et al.*, 2019). A higher analysis approach is needed to fully mine the data and this is already explored by other members of the lab.

Though inclusion of ncRNAs from Dugar *et al.* (2013)'s study may give some insights to RNA processing, since many ncRNAs were on top of genes, this was not ideal for looking at expression levels. It may be easier to not include ncRNAs that overlap genes for future analysis. Nevertheless, a few putative novel ncRNA candidates derived from our data sets have potential to be be investigated further. as_Cj1178c_1 was highly changing in a few comparisons including osmotic stress and int_Cj0747 is also highly up-regulated in osmotic stress and could possibly be a 5'UTR. *In silico* prediction of targets and/or experimental validation would be required to determine their function and behaviour across the conditions. The outcome of the analysis also highly depends on how well the genome is annotated. As many of Dugar *et al.* (2013)'s ncRNA transcripts were overlapping genes, this proved to be a drawback when analysing the data. Therefore, it may be necessary to manually curate the genome to remove such aberrations to allow for more accurate and easier interpretation of results for future work.

Possible future directions could also investigate expression of other major global regulators. In particular, Fur is known to overlap with other regulons and have cross-talk between different stresses. This data set was also used in conjunction with Chapter 4 for information on the condition-dependency of TSS and to explore ncRNAs in more detail. Nevertheless, there were some novel findings obtained from this chapter and pitfalls encountered which will help with the optimisation of future experiments.

Chapter 6

Discussion and future directions

Although not the first of its kind, this project is the first transcriptomic compendium to be carried out in *C. jejuni*. Other similar studies have investigated *S. enterica* serovar Typhimurium, *S. pneumoniae*, and *B. pseudomallei* (Kröger *et al.*, 2013; Ooi *et al.*, 2013; Slager *et al.*, 2018). Considering that *C. jejuni* is difficult to culture and maintain in the laboratory and has many unusual genetic features, the data obtained from this research project will greatly benefit the microbiology community, especially researchers working on the newly reclassified Epsilonbacteraeota.

Unfortunately, with limited time and resources it was not possible to examine many more growth and stress conditions. A useful comparison for this study would be to apply stresses at early stationary phase for a more in-depth comparison of transcriptional profiles between the growth phases. Given that growth in a microplate reader, bead-beating RNA extraction, and RNAtag-Seq were all very reproducible and variation from different sequencing runs were minimal as shown by the PCA plot (Figure 5.4), additional sequencing data can be directly compared with data from this project if the exact same protocols and methodologies are used. Thus, it may be possible to apply the same stress at early stationary phase at a later date and append this to the current data set, although the analysis would have to be repeated for normalisation. Bacteriophage and biofilm related stresses were not included in the repertoire as more time would be needed for optimisation but the transcriptional response of phage infection has been explored by Sacher *et al.* (2018). Not many different media were used in the repertoire of conditions. Given the unusual metabolic requirements of *C. jejuni*, exploring the transcriptional response in different media and supplements may uncover mysteries behind its limited metabolic requirements. Co-culture of chicken and human cells lines could also be an option for investigating the invasion of non-phagocytic cells.

It is always difficult to replicate *in vivo* conditions due to logistical limitations, thus a simpler approach was used to mimic all the growth and stress conditions *in vitro*. Perhaps a more accurate portrayal of the environment or the 'real world' would be to use a combination of stresses as bacteria do not encounter a singular stress, but multiple stresses and stimuli, often encountered simultaneously during its life. Some may argue that *in vitro* work is artificial, for example, growing bacteria for an extended period of time in

rich media and then applying nutrient starvation would be rarely encountered by bacteria. A feast-or-famine lifestyle where bacteria grow fast whilst nutrients are available, and then have prolonged periods of nutrient starvation is a more accurate portrayal (Kolter, 1993). Nevertheless, the RNA-seq data obtained from this project will give very useful insights to the regulatory networks that *C. jejuni* employs. Especially the transcriptional behaviour of many conditions which had previously not been elucidated or were carried out in different strains. A compendium of conditions allows for comparable observations and allows the use of two-factor analysis to ignore specific factors contributing to variation. In addition, specific genes can be investigated in detail across all conditions.

One important finding from this work is that osmotic stress is very unusual in its transcriptional behaviour. With the highest number of ncRNAs identified and a similar profile to 100 % chicken juice (albeit a low coverage sample) and iron repletion early stationary phase, this condition can form a basis for many future projects. Though it is unclear whether the number of unique ncRNAs found were actually due to spurious transcription. A previous microarray study by Cameron *et al.* (2012) was performed in the strain 81-176 but this was the first time the transcriptional response of osmotic stress in NCTC 11168 was sequenced.

Determining the primary transcriptome of NCTC 11168 under 22 conditions revealed many novel insights into the genome. These include a vast number of novel TSS with sigma factor preferences many of which were internal, and putative ncRNAs. The vast number of novel TSS may be attributed to the increased sensitivity using the alternative method Cappable-seq and differences in the system and media used, but more importantly it was clear that many of these novel discoveries derive from the number of different growth and stress conditions used that were missed by other studies due to the condition-dependent nature of the promoters. Although many putative novel TSS and ncRNAs were identified in this study, it is important to validate these experimentally before confirming their presence. But the identification and revelation of the presence of novel promoters and putative ncRNAs provides a foundation for experimental design rather than looking for a 'needle in a haystack'. These findings will greatly help in unravelling the regulatory networks of *C. jejuni*, which will be very useful for studying the virulence and basic biology of this pathogen and help improve preventative measures and/or treatments of Campylobacteriosis. As there were a vast number of TSS discovered, it may be logical to assume the repertoire of uncovered TSS is near saturation. There are, however, still a minority of genes which do not have an annotated promoter from published studies nor our data, although many seem to be pseudogenes and may not be transcribed, or are specific genes such as efflux pumps that may be expressed when under antibiotic selective pressure. It would be beneficial to curate a list of genes which have not yet had a promoter identified and possibly subject *C. jejuni* to the environmental changes to induce their expression, and possibly use inexpensive alternative methods such as 5' RACE to determine the TSS.

As Cappable-seq only enriches for 5' ends there is no information on the 3' end. This method has been coupled with PacBio Single Molecule, Real-Time (SMRT) sequencing i.e. SMRT-Cappable-seq to determine the transcription termination site (TTS) which would also help to confirm ncRNAs with a

defined end (Yan *et al.*, 2018). The main focus on ncRNAs have been antisense and intergenic regions. But there are also regulatory elements located in the UTRs of genes at both 5' and 3' ends. This could be an area for future exploration especially as we have the available data to examine this. Another aspect is to look for RNA-RNA interactions in order to determine the targets and function of ncRNAs, which will also validate the existence of the identified novel putative ncRNAs and this is currently ongoing in our lab.

Altogether, both RNAtag-Seq and Cappable-seq of 22 host-relevant and transmission growth and stress conditions have allowed us to determine the condition-dependent transcriptional landscape of NCTC 11168. We have achieved the aim of optimising methodologies to generate a reproducible data set that can be used as a resource in the field and for reannotation of the genome. The methods used may also be transferable to other bacterial species. Some preliminary analysis has been carried out in this project, but further work is needed to delve into the data to answer specific research questions. Overall, this study has helped unravel deeper the architecture of the NCTC 11168 genome and improved the understanding of the transcriptome in response to the changing environment. This will be beneficial for future research into the regulatory systems of *C. jejuni* and its persistence and survival as a pathogen.

Appendix

Supplementary Materials

Please visit `https://github.com/Jenna-Lam/Supplementary-material-for-thesis/blob/master/README.md` to access the supplementary materials for this thesis.

Bibliography

Achbergerová, L. and Nahálka, J. (2011). Polyphosphate - an ancient energy source and active metabolic regulator. *Microbial Cell Factories*, 10:1–14.

Acheson, D. and Allos, B. M. (2001). *Campylobacter jejuni* Infections: Update on Emerging Issues and Trends. *Clinical Infectious Diseases*, 32(8):1201–1206.

Alemka, A., Nothaft, H., Zheng, J., and Szymanski, C. M. (2013). N-Glycosylation of *Campylobacter jejuni* Surface Proteins Promotes Bacterial Fitness. *Infection and Immunity*, 81(5):1674–1682.

Andersen, M. T., Brøndsted, L., Pearson, B. M., Mulholland, F., Parker, M., Pin, C., Wells, J. M., and Ingmer, H. (2005). Diverse roles of HspR in *Campylobacter jejuni* revealed by the proteome, transcriptome and phenotypic characterization of an hspR mutant. *Microbiology*, 151(3):905–915.

Ang, C., Klerk, M., Endtz, H., Jacobs, B., Laman, J., Meché, F. G. a. V. D., and Doorn, P. a. V. (2001). Guillain-Barre ´ Syndrome- and Miller Fisher Syndrome-Associated *Campylobacter jejuni* Lipopolysaccharides Induce Anti-GM1 and Anti-GQ1b Antibodies in Rabbits. 69(4):2462–2469.

Angelini, S., My, L., and Bouveret, E. (2012). Disrupting the acyl carrier protein/SpoT interaction in vivo: Identification of ACP residues involved in the interaction and consequence on growth. *PLoS ONE*, 7(4).

Anjum, A., Brathwaite, K. J., Aidley, J., Connerton, P. L., Cummings, N. J., Parkhill, J., Connerton, I., and Bayliss, C. D. (2016). Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in Campylobacter jejuni strain NCTC11168. *Nucleic Acids Research*, 44(10):4581–4594.

Anonymous (1970). Central dogma reversed. *Nature*, 226(5252):1198–1199.

Apel, D., Ellermeier, J., Pryjma, M., DiRita, V. J., and Gaynor, E. C. (2012). Characterization of *Campylobacter jejuni* RacRS reveals roles in the heat shock response, motility, and maintenance of cell length homogeneity. *Journal of Bacteriology*, 194(9):2342–2354.

Aprianto, R., Slager, J., Holsappel, S., and Veening, J. W. (2018). High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. *Nucleic Acids Research*, 46(19):9990–10006.

Aroori, S. V., Cogan, T. A., and Humphrey, T. J. (2013). The effect of growth temperature on the pathogenicity of *Campylobacter*. *Current Microbiology*, 67(3):333–340.

Asakura, H., Hashii, N., Uema, M., Kawasaki, N., Sugita-Konishi, Y., Igimi, S., and Yamamoto, S. (2013). *Campylobacter jejuni* pdxA Affects Flagellum-Mediated Motility to Alter Host Colonization. *PLoS ONE*, 8(8).

Askoura, M., Sarvan, S., Couture, J. F., and Stintzi, A. (2016). The *Campylobacter jejuni* ferric uptake regulator promotes acid survival and cross-protection against oxidative stress. *Infection and Immunity*, 84(5):1287–1300.

Atack, J. M. and Kelly, D. J. (2009). Oxidative stress in *Campylobacter jejuni*: Responses, resistance and regulation. *Future Microbiology*, 4(6):677–690.

Atkinson, G. C., Tenson, T., and Hauryliuk, V. (2011). The RelA/SpoT Homolog (RSH) superfamily: Distribution and functional evolution of ppgpp synthetases and hydrolases across the tree of life. *PLoS ONE*, 6(8).

Avila-ramirez, C., Tinajero-trejo, M., Davidge, K. S., Monk, C. E., Kelly, D. J., and Poole, R. K. (2013). Do Globins in Microaerophilic *Campylobacter jejuni* Confer Nitrosative Stress Tolerance Under Oxygen Limitation ? 18(4):424–431.

Axelsson-Olsson, D., Svensson, L., Olofsson, J., Salomon, P., Waldenström, J., Ellström, P., and Olsen, B. (2010). Increase in acid tolerance of *Campylobacter jejuni* through coincubation with amoebae. *Applied and Environmental Microbiology*, 76(13):4194–4200.

Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pages 28–36.

Baker, E. N., Baker, H. M., and Kidd, R. D. (2002). Lactoferrin and transferrin: Functional variations on a common structural framework. *Biochemistry and Cell Biology*, 80(1):27–34.

Battesti, A. and Bouveret, E. (2006). Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. *Molecular Microbiology*, 62(4):1048–1063.

Battesti, A. and Bouveret, E. (2009). Bacteria possessing two RelA/SpoT-like proteins have evolved a specific stringent response involving The acyl carrier protein-SpoT interaction. *Journal of Bacteriology*, 191(2):616–624.

Bayliss, C. D., Bidmos, F. A., Anjum, A., Manchev, V. T., Richards, R. L., Grossier, J. P., Wooldridge, K. G., Ketley, J. M., Barrow, P. A., Jones, M. A., and Tretyakov, M. V. (2012). Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. *Nucleic Acids Research*, 40(13):5876–5889.

Behnsen, J. and Raffatellu, M. (2016). Siderophores: More than stealing iron. *mBio*, 7(6):7–9.

Birk, T., Ingmer, H., Andersen, M. T., Jørgensen, K., and Brøndsted, L. (2004). Chicken juice, a food-based model system suitable to study survival of *Campylobacter jejuni*. *Letters in Applied Microbiology*, 38(1):66–71.

Birk, T., Wik, M. T., Lametsch, R., and Knøchel, S. (2012). Acid stress response and protein induction in *Campylobacter jejuni* isolates with different acid tolerance. *BMC Microbiology*, 12.

Blomberg, P., Wagner, E. G., and Nordström, K. (1990). Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *The EMBO journal*, 9(7):2331–40.

Bloomfeld, S. J., Midwinter, A. C., Biggs, P. J., French, N. P., Marshall, J. C., Hayman, D. T., Carter, P. E., Thornley, C., Yap, R., and Benschop, J. (2018). Long-term colonization by *Campylobacter jejuni* within a human host: Evolution, antimicrobial resistance, and adaptation. *Journal of Infectious Diseases*, 217(1):103–111.

Boehm, M., Lind, J., Backert, S., and Tegtmeyer, N. (2015). *Campylobacter jejuni* serine protease HtrA plays an important role in heat tolerance, oxygen resistance, host cell adhesion, invasion, and transmigration . *European Journal of Microbiology and Immunology*, 5(1):68–80.

Boll, J. and Hendrixson, D. (2013). A Regulatory Checkpoint during Flagellar Biogenesis in *Campylobacter jejuni* Initiates Signal Transduction To Activate Transcription of Flagellar Genes. *mBio*, 4(5):1–10.

Boutte, C. C. and Crosson, S. (2008). Bacterial lifestyle shapes the regulation of stringent response activation. *Nano*, 6(9):2166–2171.

Bouwman, L. I., Niewold, P., and van Putten, J. P. (2013). Basolateral Invasion and Trafficking of *Campylobacter jejuni* in Polarized Epithelial Cells. *PLoS ONE*, 8(1).

Brás, A. M., Chatterjee, S., Wren, B. W., Newell, D. G., and Ketley, J. M. (1999). A novel *Campylobacter jejuni* two-component regulatory system important for temperature-dependent growth and colonization. *Journal of Bacteriology*, 181(10):3298–3302.

Brock, J. E., Pourshahian, S., Giliberti, J., Limbach, P. A., and Janssen, G. R. (2008). Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5-terminal AUG. *Rna*, 14(10):2159–2169.

Bronowski, C., James, C. E., and Winstanley, C. (2014). Role of environmental survival in transmission of *Campylobacter jejuni*. *FEMS Microbiology Letters*, 356(1):8–19.

Brown, H. L., Reuter, M., Salt, L. J., Cross, K. L., Betts, R. P., and Vliet, H. M. V. (2014). Chicken Juice Enhances Surface Attachment and Biofilm Formation of *Campylobacter jejuni*. 80(22):7053–7060.

Browning, D. F. and Busby, S. J. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, 14(10):638–650.

Buchanan, R. (1918). Life Phases in a Bacterial Culture. *The Journal of infectious diseases*, 23(2):109–125.

Burgess, C. M., Gianotti, A., Gruzdev, N., Holah, J., Knøchel, S., Lehner, A., Margas, E., Esser, S. S., Sela Saldinger, S., and Tresse, O. (2016). The response of foodborne pathogens to osmotic and desiccation stresses in the food chain. *International Journal of Food Microbiology*, 221:37–53.

Butcher, J., Handley, R. A., van Vliet, A. H., and Stintzi, A. (2015). Refined analysis of the *Campylobacter jejuni* iron-dependent/independent Fur- and PerR-transcriptomes. *BMC Genomics*, 16(1):1–13.

Butcher, J., Sarvan, S., Brunzelle, J. S., Couture, J.-F., and Stintzi, A. (2012). Structure and regulon of *Campylobacter jejuni* ferric uptake regulator Fur define apo-Fur regulation. *Proceedings of the National Academy of Sciences*, 109(25):10047–10052.

Butcher, J. and Stintzi, A. (2013). The transcriptional landscape of *Campylobacter jejuni* under iron replete and iron limited growth conditions. *PLoS ONE*, 8(11):1–16.

Cameron, A., Frirdich, E., Huynh, S., Parker, C. T., and Gaynor, E. C. (2012). Hyperosmotic Stress Response of *Campylobacter jejuni*. *Journal of Bacteriology*, 194(22):6116–6130.

Candon, H. L., Allan, B. J., Fraley, C. D., and Gaynor, E. C. (2007). Polyphosphate kinase 1 is a pathogenesis determinant in *Campylobacter jejuni*. *Journal of Bacteriology*, 189(22):8099–8108.

Carrillo, C. D., Taboada, E., Nash, J. H., Lanthier, P., Kelly, J., Lau, P. C., Verhulp, R., Mykytczuk, O., Sy, J., Findlay, W. A., Amoako, K., Gomis, S., Willson, P., Austin, J. W., Potter, A., Babiuk, L., Allan, B., and Szymanski, C. M. (2004). Genome-wide Expression Analyses of *Campylobacter jejuni* NCTC11168 Reveals Coordinate Regulation of Motility and Virulence by flhA. *Journal of Biological Chemistry*, 279(19):20327–20338.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28(4):464–469.

Cha, G., Chen, Z., Mo, R., Lu, G., and Gao, B. (2019). The novel regulators CheP and CheQ control the core chemotaxis operon cheVAW in *Campylobacter jejuni*. *Molecular Microbiology*, 111(1):145–158.

Chao, Y. and Vogel, J. (2010). The role of Hfq in bacterial pathogens. *Current Opinion in Microbiology*, 13(1):24–33.

Chaudhuri, R. R., Yu, L., Kanji, A., Perkins, T. T., Gardner, P. P., Choudhary, J., Maskell, D. J., and Grant, A. J. (2011). Quantitative RNA-seq analysis of the *Campylobacter jejuni* transcriptome. *Microbiology*, 157(10):2922–2932.

Cho, B. K., Kim, D., Knight, E. M., Zengler, K., and Palsson, B. O. (2014). Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: Topology and functional states. *BMC Biology*, 12:1–11.

Christensen, J. E., Pacheco, S. A., and Konkel, M. E. (2009). Identification of a *Campylobacter jejuni*-secreted protein required for maximal invasion of host cells. *Molecular Microbiology*, 73(4):650–662.

Cohn, M. T., Ingmer, H., Mulholland, F., Jørgensen, K., Wells, J. M., and Brøndsted, L. (2007). Contribution of conserved ATP-dependent proteases of *Campylobacter jejuni* to stress tolerance and virulence. *Applied and Environmental Microbiology*, 73(24):7803–7813.

Cooper, K. K., Cooper, M. A., Zuccolo, A., and Joens, L. A. (2013). Re-sequencing of a virulent strain of *Campylobacter jejuni* NCTC11168 reveals potential virulence factors. *Research in Microbiology*, 164(1):6–11.

Cremers, C. M., Knoefler, D., Vitvitsky, V., Banerjee, R., and Jakob, U. (2014). Bile salts act as effective protein-unfolding agents and instigators of disulfide stress in vivo. *Proceedings of the National Academy of Sciences*, 111(16):E1610–E1619.

Crofts, A. A., Poly, F. M., Ewing, C. P., Kuroiwa, J. M., Rimmer, J. E., Harro, C., Sack, D., Talaat, K. R., Porter, C. K., Gutierrez, R. L., DeNearing, B., Brubaker, J., Laird, R. M., Maue, A. C., Jaep, K., Alcala, A., Tribble, D. R., Riddle, M. S., Ramakrishnan, A., McCoy, A. J., Davies, B. W., Guerry, P., and Trent, M. S. (2018). *Campylobacter jejuni* transcriptional and genetic adaptation during human infection. *Nature Microbiology*, 3(4):494–502.

Cróinín, T. Ó., Backert, S., Merrell, D. S., and Mitchell, H. M. (2012). Host epithelial cell invasion by *Campylobacter jejuni*: trigger or zipper mechanism? *Frontiers in Cellular and Infection Microbiology*, 2(March):1–13.

Davis, L. and DiRita, V. (2008). Growth and laboratory maintenance of *Campylobacter jejuni*. *Current Protocols in Microbiology*, (SUPPL. 10).

Davis, L. M., Kakuda, T., and DiRita, V. J. (2009). A *Campylobacter jejuni* znuA orthologue is essential for growth in low-zinc environments and chick colonization. *Journal of Bacteriology*, 191(5):1631–1640.

Day, C. J., Semchenko, E. A., and Korolik, V. (2017). Glycoconjugates Play a Key Role in *Campylobacter jejuni* Infection : Interactions between Host and Pathogen infection : interactions between host and pathogen. 2(February 2012):1–8.

Del Rocio Leon-Kempis, M., Guccione, E., Mulholland, F., Williamson, M. P., and Kelly, D. J. (2006). The *Campylobacter jejuni* PEB1a adhesin is an aspartate/glutamate-binding protein of an ABC transporter essential for microaerobic growth on dicarboxylic amino acids. *Molecular Microbiology*, 60(5):1262–1275.

Deutscher, M. P. (2003). Degradation of Stable RNA in Bacteria. *Journal of Biological Chemistry*, 278(46):45041–45044.

Deutscher, M. P. (2015). How bacterial cells keep ribonucleases under control. *FEMS Microbiology Reviews*, 39(3):350–361.

Djordjevic, M. (2011). Redefining *Escherichia coli* σ 70 Promoter Elements: -15 Motif as a Complement of the -10 Motif. *Journal of Bacteriology*, 193(22):6305–6314.

Dugar, G., Herbig, A., Förstner, K. U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C. M. (2013). High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple *Campylobacter jejuni* Isolates. *PLoS Genetics*, 9(5).

Dugar, G., Svensson, S. L., Bischler, T., Wäldchen, S., Reinhardt, R., Sauer, M., and Sharma, C. M. (2016). The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nature Communications*, 7(May):11667.

El Abbar, F. M., Li, J., Owen, H. C., Daugherty, C. L., Fulmer, C. A., Bogacz, M., and Thompson, S. A. (2019). RNA Binding by the *Campylobacter jejuni* Post-transcriptional Regulator CsrA. *Frontiers in Microbiology*, 10(August).

Elvers, K. T., Turner, S. M., Wainwright, L. M., Marsden, G., Hinds, J., Cole, J. A., Poole, R. K., Penn, C. W., and Park, S. F. (2005). NssR, a member of the Crp-Fnr superfamily from *Campylobacter jejuni*, regulates a nitrosative stress-responsive regulon that includes both a single-domain and a truncated haemoglobin. *Molecular Microbiology*, 57(3):735–750.

Espejo, R. T. and Plaza, N. (2018). Multiple Ribosomal RNA operons in bacteria; Their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA. *Frontiers in Microbiology*, 9(JUN):1–6.

Ettwiller, L., Buswell, J., Yigit, E., and Schildkraut, I. (2016). A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics*, 17(1):1–14.

Ewing, C. P., Andreishcheva, E., and Guerry, P. (2009). Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. *Journal of Bacteriology*, 191(22):7086–7093.

Fields, J. A., Li, J., Gulbronson, C. J., Hendrixson, D. R., and Thompson, S. A. (2016). *Campylobacter jejuni* CsrA regulates metabolic and virulence associated proteins and is necessary for mouse colonization. *PLoS ONE*, 11(6):1–20.

Fields, J. A. and Thompson, S. A. (2008). *Campylobacter jejuni* CsrA mediates oxidative stress responses, biofilm formation, and host cell invasion. *Journal of Bacteriology*, 190(9):3411–3416.

Flatley, J., Barrett, J., Pullan, S. T., Hughes, M. N., Green, J., and Poolet, R. K. (2005). Transcriptional responses of *Escherichia coli* to S-nitrosoglutathione under defined chemostat conditions reveal major changes in methionine biosynthesis. *Journal of Biological Chemistry*, 280(11):10065–10072.

Flint, A., Sun, Y. Q., Butcher, J., Stahl, M., Huang, H., and Stintzi, A. (2014). Phenotypic screening of a targeted mutant library reveals *Campylobacter jejuni* defenses against oxidative stress. *Infection and Immunity*, 82(6):2266–2275.

Flint, A., Sun, Y. Q., and Stintzi, A. (2012). Cj1386 is an ankyrin-containing protein involved in heme trafficking to catalase in *Campylobacter jejuni*. *Journal of Bacteriology*, 194(2):334–345.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., and Jensen, L. J. (2013). STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(D1):808–815.

Gaca, A. O., Colomer-Winter, C., and Lemos, J. A. (2015). Many means to a common end: The intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis. *Journal of Bacteriology*, 197(7):1146–1156.

Gangaiah, D., Kassem, I. I., Liu, Z., and Rajashekara, G. (2009). Importance of polyphosphate kinase 1 for *Campylobacter jejuni* viable-but-nonculturable cell formation, natural transformation, and antimicrobial resistance. *Applied and Environmental Microbiology*, 75(24):7838–7849.

Gao, B., Vorwerk, H., Huber, C., Lara-Tejero, M., Mohr, J., Goodman, A. L., Eisenreich, W., Galán, J. E., and Hofreuter, D. (2017). Metabolic and fitness determinants for in vitro growth and intestinal colonization of the bacterial pathogen *Campylobacter jejuni*. *PLoS Biology*, 15(5):1–37.

Gardner, S. P. and Olson, J. W. (2018). Interaction of copper toxicity and oxidative stress in *Campylobacter jejuni*. *Journal of Bacteriology*, 200(21):1–11.

Gaynor, E. C., Cawthraw, S., Manning, G., MacKichan, J. K., Falkow, S., and Newell, D. G. (2004). The Genome-Sequenced Variant of *Campylobacter jejuni* NCTC 11168 and the Original Clonal Clinical Isolate Differ Markedly in Colonization, Gene Expression, and Virulence-Associated Phenotypes. *Journal of Bacteriology*, 186(2):503–517.

Gaynor, E. C., Wells, D. H., MacKichan, J. K., and Falkow, S. (2005). The *Campylobacter jejuni* stringent response controls specific stress survival and virulence-associated phenotypes. *Molecular Microbiology*, 56(1):8–27.

Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., Izard, J., Levin, J. Z., Livny, J., Earl, A. M., Gevers, D., Ward, D. V., Nusbaum, C., Birren, B. W., and Gnirke, A. (2016). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology*, 1(January):1–14.

Girardello, R., Bispo, P. J., Yamanaka, T. M., and Gales, A. C. (2012a). Cation concentration variability of four distinct Mueller-Hinton agar brands influences polymyxin B susceptibility results. *Journal of Clinical Microbiology*, 50(7):2414–2418.

Girardello, R., Bispo, P. J. M., Yamanaka, T. M., and Gales, A. C. (2012b). Cation Concentration Variability of Four Distinct Mueller-Hinton. 50(7):2414–2418.

Golden, N. J. and Acheson, D. W. K. (2002). Identification of motility and autoagglutination *Campylobacter jejuni* mutants by random transposon mutagenesis. *Infection and Immunity*, 70(4):1761–1771.

Goon, S., Kelly, J. F., Logan, S. M., Ewing, C. P., and Guerry, P. (2003). Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Molecular Microbiology*, 50(2):659–671.

Guccione, E., Del Rocio Leon-Kempis, M., Pearson, B. M., Hitchin, E., Mulholland, F., Van Diemen, P. M., Stevens, M. P., and Kelly, D. J. (2008). Amino acid-dependent growth of *Campylobacter jejuni*: Key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate. *Molecular Microbiology*, 69(1):77–93.

Guerry, P. (2007). *Campylobacter* flagella: not just for motility. *Trends in Microbiology*, 15(10):456–461.

Guerry, P., Ewing, C. P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., Majam, G., Thibault, P., and Logan, S. (2006). Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Molecular Microbiology*, 60(2):299–311.

Guerry, P., Szymanski, C. M., Prendergast, M. M., Hickey, T. E., Ewing, C. P., Pattarini, D. L., and Moran, A. P. (2002). Phase variation of *Campylobacter jejuni* 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. *Infection and Immunity*, 70(2):787–793.

Guest, M., Southgate, J., Ismail, M., Bakke, M., Poplawski, R., Connor, T. R., Loman, N. J., Thompson, S. E., Thompson, S., Bull, M. J., Kitchen, C., Smith, A., Richardson, E., Sheppard, S. K., and Pallen, M. J. (2016). CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial Genomics*, 2(9).

Gundogdu, O., da Silva, D. T., Mohammad, B., Elmi, A., Mills, D. C., Wren, B. W., and Dorrell, N. (2015). The *Campylobacter jejuni* MarR-like transcriptional regulators RrpA and RrpB both influence bacterial responses to oxidative and aerobic stresses. *Frontiers in Microbiology*, 6(JUL):1–12.

Gundogdu, O., da Silva, D. T., Mohammad, B., Elmi, A., Wren, B. W., van Vliet, A. H., and Dorrell, N. (2016). The *Campylobacter jejuni* oxidative stress regulator RrpB is associated with a genomic hypervariable region and altered oxidative stress resistance. *Frontiers in Microbiology*, 7(DEC):1–14.

Gundogdu, O., Mills, D. C., Elmi, A., Martin, M. J., Wren, B. W., and Dorrell, N. (2011). The *Campylobacter jejuni* transcriptional regulator Cj1556 plays a role in the oxidative and aerobic stress response and is important for bacterial survival in vivo. *Journal of Bacteriology*, 193(16):4238–4249.

Gursinsky, T., Gröbe, D., Schierhorn, A., Jäger, J., Andreesen, J. R., and Söhling, B. (2008). Factors and selenocysteine insertion sequence requirements for the synthesis of selenoproteins from a gram-positive anaerobe in *Escherichia coli*. *Applied and Environmental Microbiology*, 74(5):1385–1393.

Haberle, V., Forrest, A. R., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015). CAGEr: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Research*, 43(8).

Haddad, N., Burns, C. M., Bolla, J. M., Prévost, H., Fédérighi, M., Drider, D., and Cappelier, J. M. (2009). Long-term survival of *Campylobacter jejuni* at low temperatures is dependent on polynucleotide phosphorylase activity. *Applied and Environmental Microbiology*, 75(23):7310–7318.

Haddad, N., Matos, R. G., Pinto, T., Rannou, P., Cappelier, J. M., Prévost, H., and Arraiano, C. M. (2014). The RNase R from *Campylobacter jejuni* Has Unique Features and Is Involved in the First Steps of Infection. *Journal of Biological Chemistry*, 289(40):27814–27824.

Haddad, N., Saramago, M., Matos, R. G., Prevost, H., and Arraiano, C. M. (2013). Characterization of the biochemical properties of *Campylobacter jejuni* RNase III. *Bioscience Reports*, 33(6):889–901.

Haddad, N., Tresse, O., Rivoal, K., Chevret, D., Nonglaton, Q., Burns, C. M., Prévost, H., and Cappelier, J. M. (2012). Polynucleotide phosphorylase has an impact on cell biology of *Campylobacter jejuni*. *Frontiers in Cellular and Infection Microbiology*, 2(March):1–13.

Haigh, R. D. and Ketley, J. M. (2011). Growth of Campylobacter using BMG LABTECH's FLUOstar Omega equipped with ACU. pages 1–2.

Handley, R. A., Mulholland, F., Reuter, M., Ramachandran, V. K., Musk, H., Clissold, L., Le Brun, N. E., and Van Vliet, A. H. (2015). PerR controls oxidative stress defence and aerotolerance but not motility-associated phenotypes of campylobacter jejuni. *Microbiology (United Kingdom)*, 161(7):1524–1536.

Hannu, T., Mattila, L., Rautelin, H., Pelkonen, P., Lahdenne, P., Siitonen, A., and Leirisalo-Repo, M. (2002). *Campylobacter*-triggered reactive arthritis: A population-based study. *Rheumatology*, 41(3):312–318.

Hazeleger, W. C., Wouters, J. A., Rombouts, F. M., and Abee, T. (1998). Physiological activity of *Campylobacter jejuni* far below the minimal growth temperature. *Applied and Environmental Microbiology*, 64(10):3917–3922.

Hendrixson, D. R. (2006). A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. *Molecular Microbiology*, 61(6):1646–1659.

Hermans, D., Van Deun, K., Martel, A., Van Immerseel, F., Messens, W., Heyndrickx, M., Haesebrouck, F., and Pasmans, F. (2011a). Colonization factors of *Campylobacter jejuni* in the chicken gut. *Veterinary Research*, 42(1):82.

Hermans, D., Van Deun, K., Messens, W., Martel, A., Van Immerseel, F., Haesebrouck, F., Rasschaert, G., Heyndrickx, M., and Pasmans, F. (2011b). *Campylobacter* control in poultry by current intervention measures ineffective: Urgent need for intensified fundamental research. *Veterinary Microbiology*, 152(3-4):219–228.

Hermansen, G. M. M., Sazinas, P., Kofod, D., Millard, A., Andersen, P. S., and Jelsbak, L. (2018). Transcriptomic profiling of interacting nasal staphylococci species reveals global changes in gene and non-coding RNA expression. *FEMS Microbiology Letters*, 365(5):1–9.

Hitchen, P., Brzostek, J., Panico, M., Butler, J. A., Morris, H. R., Dell, A., and Linton, D. (2010). Modification of the *Campylobacter jejuni* flagellin glycan by the product of the Cj1295 homopolymeric-tract-containing gene. *Microbiology*, 156(7):1953–1962.

Hoe, C. H., Raabe, C. A., Rozhdestvensky, T. S., and Tang, T. H. (2013). Bacterial sRNAs: Regulation in stress. *International Journal of Medical Microbiology*, 303(5):217–229.

Hofreuter, D. (2014). Defining the metabolic requirements for the growth and colonization capacity of *Campylobacter jejuni*. *Frontiers in Cellular and Infection Microbiology*, 4(September):1–19.

Holmes, C. W., Penn, C. W., and Lund, P. A. (2010). The hrcA and hspR regulons of *Campylobacter jejuni*. *Microbiology*, 156(1):158–166.

Holmes, K., Mulholland, F., Pearson, B. M., Pin, C., McNicholl-Kennedy, J., Ketley, J. M., and Wells, J. M. (2005). *Campylobacter jejuni* gene expression in response to iron limitation and the role of Fur. *Microbiology*, 151(1):243–257.

Hood, M. and Skaar, E. (2012). Nutritional immunity: transition metals at the pathogen-host interface. *University of Texas, Contributions to Geology*, 10(8):361–399.

Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Molecular Cell*, pages 785–799.

Hori, H. (2014). Methylated nucleosides in tRNA and tRNA methyltransferases. *Frontiers in Genetics*, 5(MAY):1–26.

Hou, F. Q., Sun, X. T., and Wang, G. Q. (2012). Clinical manifestations of *Campylobacter jejuni* infection in adolescents and adults, and change in antibiotic resistance of the pathogen over the past 16 years. *Scandinavian Journal of Infectious Diseases*, 44(6):439–443.

Howard, S. L., Jagannathan, A., Soo, E. C., Hui, J. P., Aubry, A. J., Ahmed, I., Karlyshev, A., Kelly, J. F., Jones, M. A., Stevens, M. P., Logan, S. M., and Wren, B. W. (2009). *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infection and Immunity*, 77(6):2544–2556.

Hugdahl, M. B., Beery, J. T., and Doyle, M. P. (1988). Chemotactic behavior of *Campylobacter jejuni*. *Infection and Immunity*, 56(6):1560–1566.

Hughes, R. A., Hallett, K., Cogan, T., Enser, M., and Humphrey, T. (2009). The response of *Campylobacter jejuni* to low temperature differs from that of *Escherichia coli*. *Applied and Environmental Microbiology*, 75(19):6292–6298.

Humphrey, S., Chaloner, G., Kemmett, K., Davidson, N., Williams, N., Kipar, A., Humphrey, T., and Wigley, P. (2014). *Campylobacter jejuni* is not merely a commensal in commercial broiler chickens and affects bird welfare. *mBio*, 5(4):1–7.

Hwang, S., Jeon, B., Yun, J., and Ryu, S. (2011a). Roles of RpoN in the resistance of *Campylobacter jejuni* under various stress conditions. *BMC microbiology*, 11(1):207.

Hwang, S., Kim, M., Ryu, S., and Jeon, B. (2011b). Regulation of oxidative stress response by CosR, an essential response regulator in *Campylobacter jejuni*. *PLoS ONE*, 6(7).

Imlay, J. A. (2019). Where in the world do bacteria experience oxidative stress? *Environmental Microbiology*, 21(2):521–530.

Innocenti, N., Golumbeanu, M., D'HérouëL, A. F., Lacoux, C., Bonnin, R. A., Kennedy, S. P., Wessner, F., Serror, P., Bouloc, P., Repoila, F., and Aurell, E. (2015a). Whole-genome mapping of 5' RNA ends in bacteria by tagged sequencing: a comprehensive view in *Enterococcus faecalis*. *Rna*, 21:rna.048470.114–.

Innocenti, N., Repoila, F., and Aurell, E. (2015b). Detection and quantitative estimation of spurious double stranded DNA formation during reverse transcription in bacteria using tagRNA-seq. *RNA Biology*, (September):1067–1069.

Inoue, T., Barker, C. S., Matsunami, H., Aizawa, S. I., and Samatey, F. A. (2018). The FlaG regulator is involved in length control of the polar flagella of *Campylobacter jejuni*. *Microbiology (United Kingdom)*, 164(5):740–750.

Irving, S. E. and Corrigan, R. M. (2018). Triggering the stringent response: Signals responsible for activating (p)ppGpp synthesis in bacteria. *Microbiology (United Kingdom)*, 164(3):268–276.

Jahn, C. E., Charkowski, A. O., and Willis, D. K. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of Microbiological Methods*, 75(2):318–324.

Jaishankar, J. and Srivastava, P. (2017). Molecular basis of stationary phase survival and applications. *Frontiers in Microbiology*, 8(OCT):1–12.

Jerome, J. P., Bell, J. A., Plovanich-Jones, A. E., Barrick, J. E., Brown, C. T., and Mansfield, L. S. (2011). Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLoS ONE*, 6(1).

Johnson, J. G., Gaddy, J. A., and DiRita, V. J. (2016). The PAS domain-containing protein heur regulates heme uptake in *Campylobacter jejuni*. *mBio*, 7(6):1–11.

Johnson, T. J., Shank, J. M., and Johnson, J. G. (2017). Current and potential treatments for reducing Campylobacter colonization in animal hosts and disease in humans. *Frontiers in Microbiology*, 8(MAR):1–14.

Joslin, S. N. and Hendrixson, D. R. (2009). Activation of the *Campylobacter jejuni* FlgSR two-component system is linked to the flagellar export apparatus. *Journal of Bacteriology*, 191(8):2656–2667.

Kaakoush, N. O., Castaño-Rodríguez, N., Mitchell, H. M., and Man, S. M. (2015). Global epidemiology of Campylobacter infection. *Clinical Microbiology Reviews*, 28(3):687–720.

Kaberdina, A. C., Szaflarski, W., Nierhaus, K. H., and Moll, I. (2009). An Unexpected Type of Ribosomes Induced by Kasugamycin: A Look into Ancestral Times of Protein Synthesis? *Molecular Cell*, 33(2):227–236.

Kapranov, P. (2009). From transcription start site to cell biology. *Genome Biology*, 72(10):216.

Karki, A. B., Wells, H., and Fakhr, M. K. (2019). Retail liver juices enhance the survivability of *Campylobacter jejuni* and *Campylobacter coli* at low temperatures. *Scientific Reports*, 9(1):1–13.

Kelly, A. F., Park, S. F., Bovill, R., and Mackey, B. M. (2001). Survival of *Campylobacter jejuni* during Stationary Phase : Evidence for the Absence of a Phenotypic Stationary-Phase Response. 67(5):2248–2254.

Khanna, M. R., Bhavsar, S. P., and Kapadnis, B. P. (2006). Effect of temperature on growth and chemotactic behaviour of *Campylobacter jejuni*. *Letters in Applied Microbiology*, 43(1):84–90.

Kim, J. C., Oh, E., Kim, J., and Jeon, B. (2015). Regulation of oxidative stress resistance in *Campylobacter jejuni*, a microaerophilic foodborne pathogen. *Frontiers in Microbiology*, 6(JUL):1–12.

Kim, J. S., Artymovich, K. A., Hall, D. F., Smith, E. J., Fulton, R., Bell, J., Dybas, L., Mansfield, L. S., Tempelman, R., Wilson, D. L., and Linz, J. E. (2012). Passage of *Campylobacter jejuni* through the chicken reservoir or mice promotes phase variation in contingency genes Cj0045 and Cj0170 that strongly associates with colonization and disease in a mouse model. *Microbiology*, 158(5):1304–1316.

Klančnik, A., Botteldoorn, N., Herman, L., and Možina, S. S. (2006). Survival and stress induced expression of groEL and rpoD of *Campylobacter jejuni* from different growth phases. *International Journal of Food Microbiology*, 112(3):200–207.

Klančnik, A., Guzej, B., Jamnik, P., Vučković, D., Abram, M., and Možina, S. S. (2009). Stress response and pathogenic potential of *Campylobacter jejuni* cells exposed to starvation. *Research in Microbiology*, 160(5):345–352.

Klančnik, A., Vučković, D., Jamnik, P., Abram, M., and Možina, S. S. (2014). Stress Response and Virulence of Heat-Stressed *Campylobacter jejuni*. *Microbes and Environments*, 29(4):338–345.

Kolter, R. (1993). The Stationary Phase of the Bacterial Life Cycle. *Annual Review of Microbiology*, 47(1):855–874.

Konkel, M. E., Kim, B. J., Rivera-Amill, V., and Garvis, S. G. (1999). Bacterial secreted proteins are required for the internaliztion of *Campylobacter jejuni* into cultured mammalian cells. *Molecular microbiology*, 32(4):691–701.

Konkel, M. E., Klena, J. D., Rivera-amill, V., Monteville, M. R., Biswas, D., Raphael, B., and Mickelson, J. (2004). Secretion of Virulence Proteins from *Campylobacter jejuni* Is Dependent on a Functional Flagellar Export Apparatus. *J. Bacteriol.*, 186(11):3296–3303.

Konkel, M. E., Samuelson, D. R., Eucker, T. P., Shelden, E. A., and O'Loughlin, J. L. (2013). Invasion of epithelial cells by *Campylobacter jejuni* is independent of caveolae. *Cell Communication and Signaling*, 11(1):1–17.

Kortman, G. A., Raffatellu, M., Swinkels, D. W., and Tjalsma, H. (2014). Nutritional iron turned inside out: Intestinal stress from a gut microbial perspective. *FEMS Microbiology Reviews*, 38(6):1202–1234.

Krewulak, K. D. and Vogel, H. J. (2008). Structural biology of bacterial iron uptake. *Biochimica et Biophysica Acta - Biomembranes*, 1778(9):1781–1804.

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S. K., Hammarlöf, D. L., Canals, R., Grissom, J. E., Conway, T., Hokamp, K., and Hinton, J. C. D. (2013). An Infection-Relevant Transcriptomic Compendium for Salmonella enterica Serovar Typhimurium. *Cell Host and Microbe*, 14(6):683–695.

Kumar, A., Rajashekara, G., Gangaiah, D., and Torrelles, J. B. (2016). Polyphosphate and associated enzymes as global regulators of stress response and virulence in *Campylobacter jejuni*. *World Journal of Gastroenterology*, 22(33):7402–7414.

Kumar-Phillips, G. S., Hanning, I., and Slavik, M. (2013). Influence of acid-adaptation of *Campylobacter jejuni* on adhesion and invasion of INT 407 cells. *Foodborne Pathogens and Disease*, 10(12):1037–1043.

Kuroda, A. (2006). A polyphosphate-lon protease complex in the adaptation of *Escherichia coli* to amino acid starvation. *Bioscience, Biotechnology and Biochemistry*, 70(2):325–331.

Lahens, N. F., Kavakli, I. H., Zhang, R., Hayer, K., Black, M. B., Dueck, H., Pizarro, A., Kim, J., Irizarry, R., Thomas, R. S., Grant, G. R., and Hogenesch, J. B. (2014). IVT-seq reveals extreme bias in RNA sequencing. *Genome Biology*, 15(6):1–15.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

Lango-Scholey, L., Aidley, J., Woodacre, A., Jones, M. A., and Bayliss, C. D. (2016). High throughput method for analysis of repeat number for 28 phase variable loci of *Campylobacter jejuni* strain NCTC11168. *PLoS ONE*, 11(7):1–15.

Lastovica, A. J., On, S. L. W., and Zhang, L. (2014). The Family Campylobacteraceae. In Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E., and Thompson, F., editors, *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*, pages 307–335. Springer Berlin Heidelberg, Berlin, Heidelberg.

Le, M. T., Porcelli, I., Weight, C. M., Gaskin, D. J. H., Carding, S. R., and Vliet, a. H. M. (2012). Acid-shock of *Camyplobacter jejuni* induces flagellar gene expression and host cell invasion. *European Journal of Microbiology and Immunology*, 2(1):12–19.

Le, M. T., Veldhuizen, M. V., Porcelli, I., Bongaerts, R. J., Gaskin, D. J. H., Pearson, B. M., and Vliet, A. H. M. V. (2015). Conservation of σ 28 -Dependent Non-Coding RNA Paralogs and Predicted σ 54 -Dependent Targets in Thermophilic *Campylobacter* Species. pages 1–19.

Levy, S. E. and Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17(1):95–115.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, Z., Lou, H., Ojcius, D. M., Sun, A., Sun, D., Zhao, J., Lin, X., and Yan, J. (2014). Methyl-accepting chemotaxis proteins 3 and 4 are responsible for *Campylobacter jejuni* chemotaxis and jejuna colonization in mice in response to sodium deoxycholate. *Journal of Medical Microbiology*, 63(PART 3):343–354.

Liu, M. M., Boinett, C. J., Chan, A. C., Parkhill, J., Murphy, M. E., and Gaynor, E. C. (2018). Investigating the *Campylobacter jejuni* transcriptional response to host intestinal extracts reveals the involvement of a widely conserved iron uptake system. *mBio*, 9(4):1–18.

Liu, Y. W. and Kelly, D. J. (2015). Cytochrome c biogenesis in *Campylobacter jejuni* requires cytochrome c6 (CccA; Cj1153) to maintain apocytochrome cysteine thiols in a reduced state for haem attachment. *Molecular Microbiology*, 96(6):1298–1317.

Locher, K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature Structural and Molecular Biology*, 23(6):487–493.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21.

Luethy, P. M., Huynh, S., Parker, C. T., and Hendrixson, D. R. (2015). Analysis of the activity and regulon of the two-component regulatory system composed by Cjj81176_1484 and Cjj81176_1483 of *Campylobacter jejuni*. *Journal of Bacteriology*, 197(9):1592–1605.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K. T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE*, 7(2).

Lybecker, M., Bilusic, I., and Raghavan, R. (2014). Pervasive transcription: Detecting functional RNAs in bacteria. *Transcription*, 5(4).

Ma, J. and Betts, N. M. (2000). Zinc and Copper Intakes and Their Major Food Sources for Older Adults in the 1994–96 Continuing Survey of Food Intakes by Individuals (CSFII). *The Journal of Nutrition*, 130(11):2838–2843.

Makdessi, K., Andreesen, J. R., and Pich, A. (2001). Tungstate Uptake by a Highly Specific ABC Transporter in Eubacterium acidaminophilum. *Journal of Biological Chemistry*, 276(27):24557–24564.

Malde, A., Gangaiah, D., Chandrashekhar, K., Pina-Mimbela, R., Torrelles, J. B., and Rajashekara, G. (2014). Functional characterization of exopolyphosphatase/guanosine pentaphosphate phosphohydrolase (PPX/GPPA) of *Campylobacter jejuni*. *Virulence*, 5(4):521–533.

Maleki, F., Khosravi, A., Nasser, A., Taghinejad, H., and Azizian, M. (2016). Bacterial heat shock protein activity. *Journal of Clinical and Diagnostic Research*, 10(3):BE01–BE03.

Malik-Kale, P., Parker, C. T., and Konkel, M. E. (2008). Culture of *Campylobacter jejuni* with sodium deoxycholate induces virulence gene expression. *Journal of Bacteriology*, 190(7):2286–2297.

Mandal, R. K., Jiang, T., and Kwon, Y. M. (2017). Essential genome of *Campylobacter jejuni*. *BMC genomics*, 18(1):616.

Martin, M. (1994). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12.

Metris, A., Reuter, M., Gaskin, D. J., Baranyi, J., and van Vliet, A. H. (2011). In vivo and in silico determination of essential genes of *Campylobacter jejuni*. *BMC Genomics*, 12.

Mihaljevic, R. R., Sikic, M., Klancnik, A., Brumini, G., Mozina, S. S., and Abram, M. (2007). Environmental stress factors affecting survival and virulence of *Campylobacter jejuni*. *Microbial Pathogenesis*, 43(2):120–125.

Mikhaylina, A., Ksibe, A. Z., Scanlan, D. J., and Blindauer, C. A. (2018). Bacterial zinc uptake regulator proteins and their regulons. *Biochemical Society Transactions*, 46(4):983–1001.

Miles, A. A., Misra, S. S., and Irwin, J. O. (1938). The estimation of the bactericidal power of the blood. *The Journal of hygiene*, 38(6):732–749.

Miller, C. E., Rock, J. D., Ridley, K. A., Williams, P. H., and Ketley, J. M. (2008). Utilization of lactoferrin-bound and transferrin-bound iron by *Campylobacter jejuni*. *Journal of Bacteriology*, 190(6):1900–1911.

Miller, C. E., Williams, P. H., and Ketley, J. M. (2009). Pumping iron: Mechanisms for iron uptake by *Campylobacter*. *Microbiology*, 155(10):3157–3165.

Miravet-Verde, S., Lloréns-Rico, V., and Serrano, L. (2017). Alternative transcriptional regulation in genome-reduced bacteria. *Current Opinion in Microbiology*, 39:89–95.

Mitrophanov, A. Y. A. and Groisman, E. E. a. (2008). Signal integration in bacterial two-component regulatory systems. *Genes & development*, 22(19):2601–2611.

Mohammadpour, H., Berizi, E., Hosseinzadeh, S., Majlesi, M., and Zare, M. (2018). The prevalence of *Campylobacter* spp. in vegetables, fruits, and fresh produce: A systematic review and meta-analysis. *Gut Pathogens*, 10(1):1–12.

Moll, I., Grill, S., Gualerzi, C. O., and Bläsi, U. (2002). Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control. *Molecular Microbiology*, 43(1):239–246.

Müller, A., Beeby, M., McDowall, A. W., Chow, J., Jensen, G. J., and Clemons, W. M. (2014). Ultrastructure and complex polar architecture of the human pathogen *Campylobacter jejuni*. *MicrobiologyOpen*, 3(5):702–710.

Nachamkin, I., Yang, X. H., and Stern, N. J. (1993). Role of *Campylobacter jejuni* flagella as colonization factors for three- day-old chicks: Analysis with flagellar mutants. *Applied and Environmental Microbiology*, 59(5):1269–1273.

Nagalakshmi, U., Waern, K., and Snyder, M. (2010). RNA-seq: A method for comprehensive transcriptome analysis. *Current Protocols in Molecular Biology*, (SUPPL. 89):1–13.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Synder, M. (2008). NIH Public Access. *October*, 320(5881):1344–1349.

Naikare, H., Butcher, J., Flint, A., Xu, J., Raymond, K., and Stintzi, A. (2013). *Campylobacter jejuni* ferric-enterobactin receptor CfrA is TonB3 dependent and mediates iron acquisition from structurally different catechol siderophores. *Metallomics*, 23(1):1–7.

Naikare, H., Palyada, K., Panciera, R., Marlow, D., and Stintzi, A. (2006). Major role for FeoB in *Campylobacter jejuni* ferrous iron acquisition, gut colonization, and intracellular survival. *Infection and Immunity*, 74(10):5433–5444.

Navarro Llorens, J. M., Tormo, A., and Martínez-García, E. (2010). Stationary phase in gram-negative bacteria. *FEMS Microbiology Reviews*, 34(4):476–495.

Neal-McKinney, J. M. and Konkel, M. E. (2012). The *Campylobacter jejuni* CiaC virulence protein is secreted from the flagellum and delivered to the cytosol of host cells. *Frontiers in cellular and infection microbiology*, 2(March):31.

Negretti, N. M., Gourley, C. R., Clair, G., Adkins, J. N., and Konkel, M. E. (2017). The food-borne pathogen *Campylobacter jejuni* responds to the bile salt deoxycholate with countermeasures to reactive oxygen species. *Scientific Reports*, (October):1–11.

Ng, L. K., Sherburne, R., Taylor, D. E., and Stiles, M. E. (1985). Morphological forms and viability of *Campylobacter* species studied by electron microscopy. *Journal of Bacteriology*, 164(1):338–343.

Nguyen, H. T., Corry, J. E., and Miles, C. A. (2006). Heat resistance and mechanism of heat inactivation in thermophilic Campylobacters. *Applied and Environmental Microbiology*, 72(1):908–913.

Nguyen, T. C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F. H., Sridhar, B., Huang, N., Zhang, K., and Zhong, S. (2016). Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nature communications*, 7(May):12023.

Nichols, G. L., Richardson, J. F., Sheppard, S. K., Lane, C., and Sarran, C. (2012). *Campylobacter* epidemiology: A descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open*, 2(4):1–13.

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of Bacterial Small RNAs in Regulatory Networks. *Annual review of Biophysics*, 46:131–148.

Noinaj, N., Guillier, M., Barnard, T. J., and Buchanan, S. K. (2010). TonB-Dependent Transporters: Regulation, Structure, and Function. *Annual Review of Microbiology*, 64(1):43–60.

Nussbaumer-Pröll, A. K., Knotzer, S., Eberl, S., Reiter, B., Stimpfl, T., Jäger, W., Poschner, S., and Zeitlinger, M. (2019). Impact of erythrocytes on bacterial growth and antimicrobial activity of selected antibiotics. *European Journal of Clinical Microbiology and Infectious Diseases*, 38(3):485–495.

Ooi, W. F., Ong, C., Nandi, T., Kreisberg, J. F., Chua, H. H., Sun, G., Chen, Y., Mueller, C., Conejero, L., Eshaghi, M., Ang, R. M. L., Liu, J., Sobral, B. W., Korbsrisate, S., Gan, Y. H., Titball, R. W., Bancroft, G. J., Valade, E., and Tan, P. (2013). The Condition-Dependent Transcriptional Landscape of *Burkholderia pseudomallei*. *PLoS Genetics*, 9(9).

Palyada, K., Threadgill, D., and Stintzi, A. (2004). Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology*, 186(14):4714–4729.

Panyukov, V. V. and Ozoline, O. N. (2013). Promoters of *Escherichia coli* versus Promoter Islands: Function and Structure Comparison. *PLoS ONE*, 8(5):2–7.

Park, S. F. (2002). The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *International Journal of Food Microbiology*, 74(3):177–188.

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, a. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. a., Rajandream, M. a., Rutherford, K. M., van Vliet, a. H., Whitehead, S., and Barrell, B. G. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*, 403(6770):665–668.

Pascoe, B., Williams, L. K., Calland, J. K., Meric, G., Hitchings, M. D., Dyer, M., Ryder, J., Shaw, S., Lopes, B. S., Chintoan-Uta, C., Allan, E., Vidal, A., Fearnley, C., Everest, P., Pachebat, J. A., Cogan, T. A., Stevens, M. P., Humphrey, T. J., Wilkinson, T. S., Cody, A. J., Colles, F. M., Jolley, K. A., Maiden, M. C. J., Strachan, N., Pearson, B. M., Linton, D., Wren, B. W., Parkhill, J., Kelly, D. J., van Vliet, A. H. M., Forbes, K. J., and Sheppard, S. K. (2019). Domestication of *Campylobacter jejuni* NCTC 11168. *Microbial genomics*, 5(7).

Perez-Perez, G. I. and Kienesberger, S. (2013). *Chapter 9 – Campylobacter*. Elsevier Inc., fourth edi edition.

Petersen, L., Larsen, T. S., Ussery, D. W., On, S. L., and Krogh, A. (2003). Rpo D promoters in *Campylobacter jejuni* exhibit a strong periodic signal instead of a - 35 box. *Journal of Molecular Biology*, 326(5):1361–1372.

Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C., and Sauer, K. (2017). Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Scientific Reports*, 7(December 2016):1–15.

Phadtare, S. and Severinov, K. (2010). RNA remodeling and gene regulation by cold shock proteins. *RNA Biology*, 7(6):788–795.

Pickett, C. L., Auffenberg, T., Pesci, E. C., Sheen, V. L., and Jusuf, S. S. (1992). Iron acquisition and hemolysin production by *Campylobacter jejuni*. *Infection and Immunity*, 60(9):3872–3877.

Pielsticker, C., Glünder, G., and Rautenschlein, S. (2012). Colonization properties of *Campylobacter jejuni* in chickens. *European Journal of Microbiology and Immunology*, 2(1):61–65.

Porcelli, I., Reuter, M., Pearson, B. M., Wilhelm, T., and van Vliet, A. H. M. (2013). Parallel evolution of genome structure and transcriptional landscape in the Epsilonproteobacteria. *BMC genomics*, 14(1):616.

Priya, V. K., Sarkar, S., and Sinha, S. (2014). Evolution of tryptophan biosynthetic pathway in microbial genomes: A comparative genetic study. *Systems and Synthetic Biology*, 8(1):59–72.

Promworn, Y., Kaewprommal, P., Shaw, P. J., Intarapanich, A., Tongsima, S., and Piriyapongsa, J. (2017). ToNER: A tool for identifying nucleotide enrichment signals in feature-enriched RNAseq data. *PLoS ONE*, 12(5):1–18.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

R Core Team (2018). R: A Language and Environment for Statistical Computing.

Racki, L. R., Tocheva, E. I., Dieterle, M. G., Sullivan, M. C., Jensen, G. J., and Newman, D. K. (2017). Polyphosphate granule biogenesis is temporally and functionally tied to cell cycle exit during starvation in Pseudomonas aeruginosa . *Proceedings of the National Academy of Sciences*, 114(12):E2440–E2449.

Radomska, K. A., Ordoñez, S. R., Wösten, M. M., Wagenaar, J. A., and van Putten, J. P. (2016). Feedback control of *Campylobacter jejuni* flagellin levels through reciprocal binding of FliW to flagellin and the global regulator CsrA. *Molecular Microbiology*, 102(2):207–220.

Raines, D. J., Moroz, O. V., Blagova, E. V., Turkenburg, J. P., Wilson, K. S., and Duhme-Klair, A. K. (2016). Bacteria in an intense competition for iron: Key component of the *Campylobacter jejuni* iron uptake system scavenges enterobactin hydrolysis product. *Proceedings of the National Academy of Sciences of the United States of America*, 113(21):5850–5855.

Raphael, B. H., Pereira, S., Flom, G. A., Zhang, Q., Ketley, J. M., and Konkel, M. E. (2005). The *Campylobacter jejuni* response regulator, CbrR, modulates sodium deoxycholate resistance and chicken colonization. *Journal of Bacteriology*, 187(11):3662–3670.

Raskin, D. M., Judson, N., and Mekalanos, J. J. (2007). Regulation of the stringent response is the essential function of the conserved bacterial G protein CgtA in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(11):4636–4641.

Rasmussen, J. J., Vegge, C. S., Frøkiær, H., Howlett, R. M., Krogfelt, K. A., Kelly, D. J., and Ingmer, H. (2013). *Campylobacter jejuni* carbon starvation protein A (CstA) is involved in peptide utilization, motility and agglutination, and has a role in stimulation of dendritic cells. *Journal of Medical Microbiology*, 62(PART8):1135–1143.

Rees, D. C., Johnson, E., and Lewinson, O. (2009). ABC transporters: The power to change. *Nature Reviews Molecular Cell Biology*, 10(3):218–227.

Reeser, R. J., Medler, R. T., Billington, S. J., Jost, B. H., and Joens, L. A. (2007). Characterization of *Campylobacter jejuni* biofilms under defined growth conditions. *Applied and Environmental Microbiology*, 73(6):1908–1913.

Reid, A. N., Pandey, R., Palyada, K., Naikare, H., and Stintzi, A. (2008). Identification of *Campylobacter jejuni* genes involved in the response to acidic pH and stomach transit. *Applied and Environmental Microbiology*, 74(5):1583–1597.

Reiman, M., Laan, M., Rull, K., and Sõber, S. (2018). Effects of RNA integrity on transcript quantification by total RNA sequencing of clinically collected human placental samples. *The FASEB Journal*, 16:3298–3331.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., and Bader, G. D. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, 14(2):482–517.

Reuter, M., Mallett, A., Pearson, B. M., and Van Vliet, A. H. M. (2010). Biofilm formation by *Campylobacter jejuni* is increased under aerobic conditions. *Applied and Environmental Microbiology*, 76(7):2122–2128.

Reuter, M., Periago, P. M., Mulholland, F., Brown, H. L., and van Vliet, A. H. (2015). A PAS domain-containing regulator controls flagella-flagella interactions in *Campylobacter jejuni*. *Frontiers in Microbiology*, 6(JUL):1–13.

Richard, K. L., Kelley, B. R., and Johnson, J. G. (2019). Heme uptake and utilization by gram-negative bacterial pathogens. *Frontiers in Cellular and Infection Microbiology*, 9(MAR):1–12.

Riddle, M. S. and Guerry, P. (2016). Status of vaccine research and development for *Campylobacter jejuni*. *Vaccine*, 34(26):2903–2906.

Ridley, K. A., Rock, J. D., Li, Y., and Ketley, J. M. (2006). Heme utilization in *Campylobacter jejuni*. *Journal of Bacteriology*, 188(22):7862–7875.

Rivera-Amill, V., Kim, B., Seshu, J., and Konkel, M. (2001). Secretion of the Virulence-Associated *Campylobacter* Invasion Antigens from *Campylobacter jejuni* Requires a Stimulatory Signal . *The Journal of Infectious Diseases*, 183(11):1607–1616.

Robinson, D. (1981). Infective dose of *Campylobacter jejuni* in milk. *British Medical Journal*, 282(May):1981.

Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D., Alston, M., Stringer, M. F., Betts, R. P., Baranyi, J., Peck, M. W., and Hinton, J. C. (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology*, 194(3):686–701.

Rollins, D. M. and Colwell, R. R. (1986). Viable but nonculturable stage of *Campylobacter jejuni* and its role in survival in the natural aquatic environment. *Applied and Environmental Microbiology*, 52(3):531–538.

Ronneau, S. and Hallez, R. (2019). Make and break the alarmone: regulation of (p)ppGpp synthetase/hydrolase enzymes in bacteria. *FEMS Microbiology Reviews*, 43(4):389–400.

Roure, S., Bonis, M., Chaput, C., Ecobichon, C., Mattox, A., Barrière, C., Geldmacher, N., Guadagnini, S., Schmitt, C., Prévost, M. C., Labigne, A., Backert, S., Ferrero, R. L., and Boneca, I. G. (2012). Peptidoglycan maturation enzymes affect flagellar functionality in bacteria. *Molecular Microbiology*, 86(4):845–856.

Rubinchik, S., Seddon, A., and Karlyshev, A. V. (2012). Molecular mechanisms and biological role of *Campylobacter jejuni* attachment to host cells . *European Journal of Microbiology and Immunology*, 2(1):32–40.

Rudat, A. K., Pokhrel, A., Green, T. J., and Gray, M. J. (2018). Mutations in *Escherichia coli* polyphosphate kinase that lead to dramatically increased in vivo polyphosphate levels. *Journal of Bacteriology*, 200(6):1–20.

Sacher, J. C., Flint, A., Butcher, J., Blasdel, B., Reynolds, H. M., Lavigne, R., Stintzi, A., and Szymanski, C. M. (2018). Transcriptomic analysis of the *Campylobacter jejuni* response to T4-like phage NCTC 12673 infection. *Viruses*, 10(6).

Schlüter, J.-P., Reinkensmeier, J., Barnett, M. J., Lang, C., Krol, E., Giegerich, R., Long, S. R., and Becker, A. (2013). Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics*, 14(1):156.

Schroeder, L., Lazarevskiy, P., and Ikui, A. (2018). Tryptophan recovers sensitivity to cell membrane stress in Saccharomyces cerevisiae. *bioRxiv*, 94(2):19–22.

Sébald, M. and Véron, M. (1963). BASE DNA CONTENT AND CLASSIFICATION OF VIBRIOS. *Ann Inst Pasteur (Paris).*, (105):897–910.

Sellars, M. J., Hall, S. J., and Kelly, D. J. (2002). Growth of *Campylobacter jejuni* Supported by Respiration of Fumarate, Nitrate, Nitrite, Trimethylamine-N-Oxide, or Dimethyl Sulfoxide Requires Oxygen. 184(15):4187–4196.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F., and Vogel, J. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, 464(7286):250–255.

Sharma, C. M. and Vogel, J. (2014). Differential RNA-seq: The approach behind and the biological insight gained. *Current Opinion in Microbiology*, 19(1):97–105.

Shcherbakov, D. V. and Garber, M. B. (2000). Overlapping genes in bacterial and phage genomes. *Molekulyarnaya Biologiya*, 34(4):572–583.

Sheldon, J. R., Laakso, H. A., and Heinrichs, D. E. (2015). Iron Acquisition Strategies of Bacterial Pathogens. *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition*, 4(2):43–85.

Sheppard, S. K., Dallas, J. F., Strachan, N. J. C., MacRae, M., McCarthy, N. D., Wilson, D. J., Gormley, F. J., Falush, D., Ogden, I. D., Maiden, M. C. J., and Forbes, K. J. (2009). *Campylobacter* genotyping to determine the source of human infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 48(8):1072–1078.

Shishkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R. P., Rudy, R. F., Patel, M. M., Novod, N., Hung, D. T., Gnirke, A., Garber, M., Guttman, M., and Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature method*, 12(4):323–5.

Singh, S. S., Singh, N., Bonocora, R. P., Fitzgerald, D. M., Wade, J. T., and Grainger, D. C. (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes and Development*, 28(3):214–219.

Skarp, C. P., Hänninen, M. L., and Rautelin, H. I. (2016). Campylobacteriosis: The role of poultry meat. *Clinical Microbiology and Infection*, 22(2):103–109.

Skirrow, M. B. (1977). *Campylobacter* enteritis: a "new" disease. *British medical journal*, 2(6078):9–11.

Skirrow, M. B. and Benjamin, J. (1980). '1001' *Campylobacters*: Cultural characteristics of intestinal campylobacters from man and animals. *Journal of Hygiene*, 85(3):427–442.

Slager, J., Aprianto, R., and Veening, J.-W. (2018). Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. *Nucleic Acids Research*, pages 1–19.

Smart, J. P., Cliff, M. J., and Kelly, D. J. (2009). A role for tungsten in the biology of *Campylobacter jejuni*: Tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter. *Molecular Microbiology*, 74(3):742–757.

Smedley, P. L. and Kinniburgh, D. G. (2017). Molybdenum in natural waters: A review of occurrence, distributions and controls. *Applied Geochemistry*, 84:387–432.

Smith, C. K., AbuOun, M., Cawthraw, S. A., Humphrey, T. J., Rothwell, L., Kaiser, P., Barrow, P. A., and Jones, M. A. (2008). *Campylobacter* colonization of the chicken induces a proinflammatory response in mucosal tissues. *FEMS Immunology and Medical Microbiology*, 54(1):114–121.

Sproston, E. L., Wimalarathna, H. M. L., and Sheppard, S. K. (2018). Trends in fluoroquinolone resistance in *Campylobacter*. *Microbial Genomics*, 4(8).

Stahl, M., Butcher, J., and Stintzi, A. (2012). Nutrient Acquisition and Metabolism by *Campylobacter jejuni*. *Frontiers in Cellular and Infection Microbiology*, 2(February):1–10.

Stahl, M., Li, J., Stintzi, A., Friis, L. M., Liu, X., Szymanski, C. M., and Nothaft, H. (2011). L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proceedings of the National Academy of Sciences*, 108(17):7194–7199.

Starosta, A. L., Lassak, J., Jung, K., and Wilson, D. N. (2014). The bacterial translation stress response. *FEMS Microbiology Reviews*, 38(6):1172–1201.

Stintzi, A. (2002). Gene expression profile of *Campylobacter jejuni* in response to Growth Temperature Variation. *Journal of Bacteriology*, 185(6):2009–2016.

Stoakes, E. A. (2017). *Campylobacter jejuni : Deciphering the role of FlhF in flagellar biogenesis*. PhD thesis, University of Warwick.

Stougaard, P., Molin, S., and Nordstrom, K. (1981). RNAs involved in copy-number control and incompatibility of plasmid R1. *Proceedings of the National Academy of Sciences*, 78(10):6008–6012.

Strachan, N. J. C., Rotariu, O., Cowden, J., and Sheppard, S. K. (2014). Europe PMC Funders Group Identifying the seasonal origins of human campylobacteriosis. 141(6):1267–1275.

Svensson, S., Frirdich, E., and Gaynor, E. C. (2008). Survival Strategies of *Campylobacter jejuni*: Stress Responses, the Viable but Nonculturable State, and Biofilms. In Nachamin, I., Szymanski, C. M., and Blaser, M., editors, *Campylobacter, Third Edition*, chapter 32, pages 571–590. ASM Press, 3rd edition.

Svensson, S. L., Davis, L. M., Mackichan, J. K., Allan, B. J., Pajaniappan, M., Thompson, S. A., and Gaynor, E. C. (2009). The CprS sensor kinase of the zoonotic pathogen *Campylobacter jejuni* influences biofilm formation and is required for optimal chick colonization. 71(1):253–272.

Svensson, S. L., Hyunh, S., Parker, C. T., and Gaynor, E. C. (2015). The *Campylobacter jejuni* CprRS two-component regulatory system regulates aspects of the cell envelope. *Molecular Microbiology*, 96(1):189–209.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(Database issue):447–52.

Szymanski, C. M., Logan, S. M., Linton, D., and Wren, B. W. (2003). *Campylobacter* - A tale of two protein glycosylation systems. *Trends in Microbiology*, 11(5):233–238.

Taema, M. M., Bull, J. C., Macgregor, S. K., Flach, E. J., Boardman, W. S., and Routh, A. D. (2008). Retrospective study of *Campylobacter* infection in a zoological collection. *Applied and Environmental Microbiology*, 74(5):1332–1338.

Taheri, N., Fällman, M., Wai, S. N., and Fahlgren, A. (2019). Accumulation of virulence-associated proteins in *Campylobacter jejuni* Outer Membrane Vesicles at human body temperature. *Journal of Proteomics*, 195(January):33–40.

Taheri, N., Mahmud, A. K., Sandblad, L., Fällman, M., Wai, S. N., and Fahlgren, A. (2018). *Campylobacter jejuni* bile exposure influences outer membrane vesicles protein content and bacterial interaction with epithelial cells. *Scientific Reports*, 8(1):1–14.

Taveirne, M. E., Sikes, M. L., and Olson, J. W. (2009). Molybdenum and tungsten in *Campylobacter jejuni*: Their physiological role and identification of separate transporters regulated by a single ModE-like protein. *Molecular Microbiology*, 74(3):758–771.

Taveirne, M. E., Theriot, C. M., Livny, J., and DiRita, V. J. (2013). The Complete *Campylobacter jejuni* Transcriptome during Colonization of a Natural Host Determined by RNAseq. *PLoS ONE*, 8(8):1–18.

Taylor, P., Stahl, M., and Vallance, B. A. (2015). Insights into *Campylobacter jejuni* colonization of the mammalian intestinal tract using a novel mouse model of infection Martin. *Gut microbes*, (April 2015):37–41.

Teunis, P. F., Bonačić Marinović, A., Tribble, D. R., Porter, C. K., and Swart, A. (2018). Acute illness from *Campylobacter jejuni* may require high doses while infection occurs at low doses. *Epidemics*, 24(January):1–20.

Tom-Yew, S. A. L., Cui, D. T., Bekker, E. G., and Murphy, M. E. P. (2005). Anion-independent iron coordination by the *Campylobacter jejuni* ferric binding protein. *Journal of Biological Chemistry*, 280(10):9283–9290.

Tran, P., Leclerc, D., Chan, M., Pai, A., Hiou-Tim, F., Wu, Q., Goyette, P., Artigas, C., Milos, R., and Rozen, R. (2002). Multiple transcription start sites and alternative splicing in the methylenetetrahydrofolate reductase gene result in two enzyme isoforms. *Mammalian Genome*, 13(9):483–492.

Tribble, D. R., Baqar, S., Scott, D. A., Oplinger, M. L., Trespalacios, F., Rollins, D., Walker, R. I., Clements, J. D., Walz, S., Gibbs, P., Burg, E. F., Moran, A. P., Applebee, L., and Bourgeois, A. L. (2010). Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans. *Infection and Immunity*, 78(4):1750–1759.

Turonova, H., Haddad, N., Hernould, M., Chevret, D., Pazlarova, J., and Tresse, O. (2017). Profiling of *Campylobacter jejuni* proteome in exponential and stationary phase of growth. *Frontiers in Microbiology*, 8(MAY):1–12.

Ugarte-Ruiz, M., Domínguez, L., Corcionivoschi, N., Wren, B. W., Dorrell, N., and Gundogdu, O. (2018). Exploring the oxidative, antimicrobial and genomic properties of *Campylobacter jejuni* strains isolated from poultry. *Research in Veterinary Science*, 119(October 2017):170–175.

Urdaneta, V. and Casadesús, J. (2017). Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts. *Frontiers in Medicine*, 4(October):1–13.

van der Hooft, J. J., Alghefari, W., Watson, E., Everest, P., Morton, F. R., Burgess, K. E., and Smith, D. G. (2018). Unexpected differential metabolic responses of *Campylobacter jejuni* to the abundant presence of glutamate and fucose. *Metabolomics*, 14(11):1–15.

van Vliet, A. H., Ketley, J. M., Park, S. F., and Penn, C. W. (2002). The role of iron in *Campylobacter* gene regulation, metabolism and oxidative stress defense. *FEMS Microbiology Reviews*, 26.

van Vliet, A. H. M., Stintzi, A., and Ketley, J. M. (2008). Iron Metabolism, Transport, and Regulation. In *Campylobacter, Third Edition*, pages 591–610. American Society of Microbiology.

Varsaki, A., Murphy, C., Barczynska, A., Jordan, K., and Carroll, C. (2015). The acid adaptive tolerance response in *Campylobacter jejuni* induces a global response, as suggested by proteomics and microarrays. *Microbial Biotechnology*, 8(6):974–988.

Vegge, C. S., Jansen van Rensburg, M. J., Rasmussen, J. J., Maiden, M. C., Johnsen, L. G., Danielsen, M., MacIntyre, S., Ingmer, H., and Kelly, D. J. (2016). Glucose metabolism via the entner-doudoroff pathway in *Campylobacter*: A rare trait that enhances survival and promotes biofilm formation in some isolates. *Frontiers in Microbiology*, 7(NOV).

Véron, M. and Chatelain, R. (1973). Taxonomic Study of the Genus *Campylobacter* Sebald and Veron and Designation of the Neotype Strain for the Type Species, *Campylobacter fetus* (Smith and Taylor) Sebald and Veron. *International Journal of Systematic Bacteriology*, 23(2):122–134.

Vorwerk, H., Mohr, J., Huber, C., Wensel, O., Schmidt-Hohagen, K., Gripp, E., Josenhans, C., Schomburg, D., Eisenreich, W., and Hofreuter, D. (2014). Utilization of host-derived cysteine-containing peptides overcomes the restricted sulphur metabolism of *Campylobacter jejuni*. *Molecular Microbiology*, 93(6):1224–1245.

Wade, J. T. and Grainger, D. C. (2014). Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Publishing Group*, 12(9):647–653.

Wade, J. T. and Grainger, D. C. (2018). Spurious transcription and its impact on cell function. *Transcription*, 9(3):182–189.

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131(4):281–285.

Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., Sievert, S. M., Simon, J., Campbell, B. J., Hanson, T. E., Woyke, T., Klotz, M. G., and Hugenholtz, P. (2018). Addendum: Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteraeota (phyl. nov.) [Front. Microbiol., 8, (2017), (682)] doi: 10.3389/fmicb.2017.00682. *Frontiers in Microbiology*, 9(APR).

Watson, R. and Galán, J. (2007). *Campylobacter jejuni* Survives within Epithelial Cells by Avoiding Delivery to Lysosomes. *Plos Pathogens*, 3(2):e23.

Weingarten, R. A., Grimes, J. L., and Olson, J. W. (2008). Role of *Campylobacter jejuni* respiratory oxidases and reductases in host colonization. *Applied and Environmental Microbiology*, 74(5):1367–1375.

WHO (1980). Enteric infections due to *Campylobacter*, *Yersinia*, *Salmonella*, and *Shigella*. *Bulletin of the World Health Organization*, 58(4):519–537.

WHO (2018). *Campylobacter*.

Wood, J. M. (2015). Bacterial responses to osmotic challenges. *The Journal of General Physiology*, 145(5):381–388.

Workman, S. N., Mathison, G. E., Marc, C., and Lavoie, M. C. (2005). Pet Dogs and Chicken Meat as Reservoirs of Campylobacter spp . in Barbados Pet Dogs and Chicken Meat as Reservoirs of Campylobacter spp . in Barbados. *Journal of clinical microbiology*, 43(6):2642–2650.

Wösten, M. M., Van Dijk, L., Veenendaal, A. K., De Zoete, M. R., Bleumink-Pluijm, N. M., and Van Putten, J. P. (2010a). Temperature-dependent FlgM/FliA complex formation regulates *Campylobacter jejuni* flagella length. *Molecular Microbiology*, 75(6):1577–1591.

Wösten, M. M. S. M., Parker, C. T., Van Mourik, A., Guilhabert, M. R., Van Dijk, L., and Van Putten, J. P. M. (2006). The *Campylobacter jejuni* PhosS/PhosR operon represents a non-classical phosphate-sensitive two-component system. *Molecular Microbiology*, 62(1):278–291.

Wösten, M. M. S. M., Van Dijk, L., Parker, C. T., Guilhabert, M. R., Van Der Meer-Janssen, Y. P. M., Wagenaar, J. A., and Van Putten, J. P. M. (2010b). Growth phase-dependent activation of the DccRS regulon of *Campylobacter jejuni*. *Journal of Bacteriology*, 192(11):2729–2736.

Wright, J. A., Grant, A. J., Hurd, D., Harrison, M., Guccione, E. J., Kelly, D. J., and Maskell, D. J. (2009). Metabolite and transcriptome analysis of *Campylobacter jejuni* in vitro growth reveals a stationary-phase physiological switch. *Microbiology*, 155(1):80–94.

Wurtzel, O., Sesto, N., Mellin, J. R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., and Sorek, R. (2012). Comparative transcriptomics of pathogenic and non-pathogenic Listeria species. *Molecular Systems Biology*, 8(1):1–14.

Yan, B., Boitano, M., Clark, T. A., and Ettwiller, L. (2018). SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nature Communications*, 9(3676).

Yang, X., Chockalingam, S. P., and Aluru, S. (2013). A survey of error-correction methods for next-generation sequencing. *Briefings in Bioinformatics*, 14(1):56–66.

Yao, R., Burr, D. H., Doig, P., Trust, T. J., Niu, H., and Guerry, P. (1994). Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Molecular Microbiology*, 14(5):883–893.

Zautner, A. E., Malik Tareen, A., Groß, U., and Lugert, R. (2012). Chemotaxis in *Campylobacter jejuni*. *European Journal of Microbiology and Immunology*, 2(1):24–31.

Zeng, X., Xu, F., and Lin, J. (2013). Specific TonB-ExbB-ExbD energy transduction systems required for ferric enterobactin acquisition in *Campylobacter*. *FEMS Microbiology Letters*, 347(1):83–91.

Zhuang, H., Bowker, B. C., Jeff Buhr, R., Bourassa, D. V., and Kiepper, B. H. (2013). Effects of broiler carcass scalding and chilling methods on quality of early-deboned breast fillets. *Poultry Science*, 92(5):1393–1399.

Zumbo, P. (1979). Phenol - chloroform Extraction. pages 1–7.